



(12) 发明专利

(10) 授权公告号 CN 102286456 B

(45) 授权公告日 2014. 04. 30

(21) 申请号 201110177668. X

(22) 申请日 2006. 03. 31

(30) 优先权数据

05075781. 4 2005. 04. 04 EP

60/669, 243 2005. 04. 07 US

(62) 分案原申请数据

200680011307. 5 2006. 03. 31

(73) 专利权人 拜尔作物科学公司

地址 比利时迪海姆

(72) 发明人 K·达伦 R·鲁伊特

(74) 专利代理机构 北京市中咨律师事务所

11247

代理人 胡志君 黄革生

(51) Int. Cl.

C12N 15/09 (2006. 01)

C12N 5/10 (2006. 01)

C12N 15/63 (2006. 01)

A01H 5/00 (2006. 01)

A01H 1/02 (2006. 01)

(56) 对比文件

WO 03004659 A3, 2003. 09. 25,

卞广兴和郭葆玉. DNA 双链断裂与同源重

组修复机理研究进展. 《国外医学分子生物学分册》. 2001, 第 23 卷 (第 2 期), 75-77.

吕慧娟和吕祖坝. 转基因植物中标记基因的剔除. 《植物生理与分子生物学学报》. 2004, 第 30 卷 (第 2 期), 121-126.

Holger Puchta. Marker-free transgenic plants. 《Plant Cell, Tissue and Organ Culture》. 2003, 第 74 卷 (第 2 期), 123-134.

Barbara Hohn. Elimination of selection markers from transgenic plants. 《Current Opinion in Biotechnology》. 2001, 第 12 卷 (第 2 期), 139-143.

审查员 高巍

权利要求书2页 说明书21页

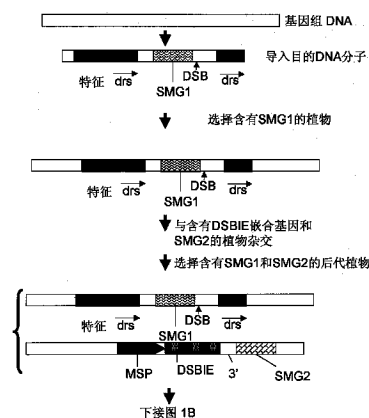
序列表22页 附图6页

(54) 发明名称

用于去除所选择 DNA 序列的方法和手段

(57) 摘要

本发明描述用于使用处于小孢子特异性启动子控制下表达切点罕见双链断裂诱导 DNA 内切核酸酶通过在两个同向重复 DNA 序列间的染色体重组而从 DNA 分子内精确去除所选择亚片段的方法。本方法可以应用于用来将植物细胞和植物内的靶 DNA 片段精确交换成目的 DNA 片段的方法内。



1. 用于将植物基因组内的靶 DNA 序列交换成目的 DNA 序列的方法,包括如下步骤:
 - a. 通过在植物细胞中导入第一个双链 DNA 断裂诱导酶在植物细胞基因组内预先选择位点处诱导第一个双链 DNA 断裂,所述预先选择位点位于所述靶 DNA 序列内,所述双链 DNA 断裂诱导酶的缩写是 DSBI 酶;
 - b. 将目的 DNA 分子导入所述的植物细胞,所述的 DNA 分子包含:
 - i. 位于两个侧翼 DNA 区域之间的所述目的 DNA 序列,其中所述的侧翼 DNA 区域与所述植物细胞基因组内所述靶 DNA 序列的侧翼 DNA 区域和所述预先选择位点的侧翼 DNA 区域具有至少 80% 序列同源性;
 - ii. 位于所述侧翼 DNA 区域之间的选择标记基因或筛选标记基因,所述选择标记基因或筛选标记基因还位于侧翼 DNA 区域之一与以同向重复方式存在的至少部分的所述侧翼 DNA 区域之一的另一拷贝之间,其中所述另一拷贝标示为部分侧翼 DNA 区域;
 - iii. 位于侧翼 DNA 区域之一与以同向重复方式存在的所述部分侧翼 DNA 区域之间的 DSBI 酶的识别位点;
 - c. 选择包含所述选择标记或筛选标记的植物细胞的群体;
 - d. 选择在其中所述选择标记或筛选标记已经借助所述侧翼 DNA 区域通过同源重组而导入的植物细胞并从所述植物细胞再生植物;
 - e. 将包含所述选择标记基因的所述再生植物或其后代植物与包含编码 DSBI 酶的嵌合基因的植物杂交,所述嵌合基因包含如下有效连接的 DNA 区段:
 - iv. 小孢子特异性启动子;
 - v. 编码可识别位于所述目的 DNA 分子内的所述识别位点的双链 DNA 断裂诱导酶的 DNA 区域;
 - vi. 转录终止和聚腺苷酸化区域;
 - f. 选择包含所述选择标记基因或筛选标记基因和编码 DSBI 酶的所述嵌合基因的后代植物;
 - g. 将所述的后代植物与另一植物杂交,由此使用所述后代植物作为花粉供体;
 - h. 选择包含编码 DSBI 酶的所述嵌合基因的后代植物的群体;和
 - i. 选择如此的后代植物,在所述后代植物中所述的选择标记基因或筛选标记基因通过在所述侧翼 DNA 区域之一与包含所述侧翼 DNA 区域之一的一部分的部分侧翼 DNA 区域之间的同源重组被缺失。
2. 权利要求 1 所述的方法,其中所述的第一个 DSBI 酶不识别位于所述目的 DNA 分子内的 DSBI 酶的所述识别位点。
3. 权利要求 2 所述的方法,其中所述的第一个 DSBI 酶和识别位于所述目的 DNA 分子内的所述识别位点的所述 DSBI 酶是选自 I-Sce I、I-Chu I、I-Dmo I、I-Cre I、I-Csm I、I-Ceu I、I-Sce II、I-Sce III、HO、PI-Civ I、PI-Ctr I、PI-Aae I、PI-BSU I、PI-DhaI、PI-Dra I、PI-Mav I、PI-Mch I、PI-Mfu I、PI-Mfl I、PI-Mga I、PI-Mgo I、PI-Min I、PI-Mka I、PI-Mle I、PI-Mma I、PI-Msh I、PI-Msm I、PI-Mth I、PI-Mtu I、PI-Mxe I、PI-Npu I、PI-Pfu I、PI-Rma I、PI-Spb I、PI-Ssp I、PI-Fac I、PI-Mja I、PI-Pho I、PI-Tag I、PI-Thy I 或 PI-Tsp I 或包含 Zn 指 DNA 结合结构域和 DNA 切割结构域的嵌合内切核酸酶的不同 DSBI 酶。

4. 权利要求 1 至 3 中任一项所述的方法,其中识别位于所述目的 DNA 分子内的 DSB1 酶的所述识别位点的所述 DSB1 酶是 I-SceI。

5. 权利要求 4 所述的方法,其中编码所述双链 DNA 断裂诱导酶的所述 DNA 区域由 SEQ ID No1 或 SEQ ID No2 的核苷酸序列组成。

6. 权利要求 1 至 3 中任一项所述的方法,其中所述小孢子特异性启动子是由 SEQ ID No3 的核苷酸序列组成的启动子。

7. 权利要求 1 至 3 中任一项所述的方法,其中编码 DSB1 酶的所述嵌合基因由 SEQ ID No. 6 中从核苷酸 1941 至 3913 的核苷酸序列组成。

8. 用于产生植物的方法,所述植物是植物基因组内的靶 DNA 序列替换成目的 DNA 序列的植物,所述方法包括实施权利要求 1 至 7 中任一项所述的方法。

9. 用于通过在靶序列内的预先选择位点处诱导双链断裂而将植物细胞基因组内的靶 DNA 序列替换成目的 DNA 序列的 DNA 载体,所述 DNA 载体包含:

a. 位于两个侧翼 DNA 区域之间的所述目的 DNA 序列,其中所述的侧翼 DNA 区域与所述靶 DNA 序列的侧翼 DNA 区域和所述预先选择位点的侧翼 DNA 区域具有至少 80% 序列同源性;

b. 位于所述侧翼 DNA 区域之间的选择标记基因或筛选标记基因,所述选择标记基因或筛选标记基因还位于侧翼 DNA 区域之一与以同向重复方式存在的包含所述侧翼 DNA 区域之一的一部分的部分侧翼 DNA 区域之间;和

c. 位于所述侧翼 DNA 区域之一与以同向重复方式存在的所述部分侧翼 DNA 区域之间的 DSB1 酶的识别位点。

用于去除所选择 DNA 序列的方法和手段

[0001] 本申请是中国专利申请 200680011307.5 的分案申请,原申请的申请日是 2006 年 3 月 31 日,发明名称是“用于去除所选择 DNA 序列的方法和手段”。

发明领域

[0002] 本发明涉及在去除步骤期间不求助于体外 (in vitro) 培养,允许高效去除事先已导入所述植物内的目的 DNA 序列中所选择部分例如选择标记基因或筛选标记基因的方法和手段。去除方法可以作为用于在植物细胞和植物中将靶 DNA 序列经同源重组精确交换为目的 DNA 序列的方法的一部分使用,由此随后可以去除同源重组阶段期间为临时选择基因替换事件所用的选择标记或筛选标记,而不会留下痕迹并且在去除步骤期间不求助于体外培养。

技术背景

[0003] 去除导入植物细胞或植物内,但在导入后随即因多种原因变成无用或甚至是多余的外源 DNA 中所选择亚片段已经是热点研究的主题。此类序列的实例例如是需要用于分离转基因植物而在成熟植物内不再需要的选择标记基因。实现高效消除选择标记基因的方法主要依赖位点特异性重组或转座(见例如 Hohn 等,Plant BioTechnology 第 139-143 页)。

[0004] Siebert 和 Puchta(2002) 描述可以通过同源重组和通过非同源末端连接从高等真核生物的基因组内高效地切除侧翼为切点罕见限制酶位点的转基因序列。

[0005] W003/004659 涉及重组系统并涉及用于从真核生物的染色体 DNA 中去除核酸序列的方法。该文件还涉及含有所述系统或通过该方法产生的转基因生物(优选地是植物)。

[0006] 然而该方法基本上需要利用体外培养方法来鉴定或选择其中待去除 DNA 序列的缺失已经有效发生的那些植物细胞并从此类细胞再生植物。

[0007] 美国专利申请 2005/0060769 提出从第一个转基因玉米 (Zea mays) 植物细胞中制备重组的转基因玉米植物或植物细胞的方法,其中相对于第一个转基因植物细胞内的转基因的遗传结构,重组植物或植物细胞内的转基因由于同源重组介导性转基因缺失作用而具有改变的遗传结构。

[0008] 在下文,包括在权利要求书内,描述如此方法和手段的不同实施方案,该方法和手段可不求助于体外培养方法,高效去除已事先导入植物细胞内的 DNA 分子的部分中所选择的亚序列。

[0009] W097/30166 或美国专利 6,407,314 描述可以用于在小孢子内表达基因的来自烟草的小孢子特异性基因的启动子片段。

[0010] 由本发明解决的另一个问题涉及定向和精确地将植物细胞内的靶 DNA 序列通过同源重组交换成替换性 DNA 序列,而不留下方法的痕迹,并且在同源重组的初始步骤后不求助于体外培养方法。为此目的,可以方便地施用如此方法,即其可以经染色体同源重组而高效地去除事先已插入基因组、优选地是植物细胞的核基因组内 DNA 分子的部分中所选择的亚序列。

[0011] 早已认识到需要对植物内转基因整合的位点进行控制,并且已经开发数种方法试图满足这种需要(综述见 Kumar 和 Fladung, 2001, Trends in Plant Science, 6, pp155-159)。这些方法基本上依赖于以同源重组为基础的转基因整合,该策略已经成功地应用于原核生物和低等真核生物(见例如 EP0317509 或相应的出版物 Paszkowski 等, 1988, EMBO J., 7, 第 4021-4026 页)。然而对于植物,用于转基因整合的主导机制以很少涉及重组 DNA 链间同源性的异常重组为基础。因而,此领域内的主要挑战是检测由所导入外源 DNA 经异常重组的极高效整合所掩盖的稀有同源重组事件。

[0012] 解决此问题的一种方式是通过异常重组发生的整合事件,如 W094/17176 内举例。

[0013] 解决此问题的另一种方式是通过切点罕见内切核酸酶如 I-SceI 诱导双链 DNA 断裂而激活靶基因座。已经使用农杆菌 (Agrobacteria) 递送修复性 DNA 至植物细胞证实这种技术增加同源重组频率至少 2 个数量级 (Puchta 等, 1996, Proc. Natl. Acad. Sci. U. S. A., 93, 第 5055-5060 页)。

[0014] W096/14408 描述编码 I-SceI 酶的分离的 DNA。该 DNA 序列可以并入克隆载体及表达载体、转化的细胞系和转基因动物。该载体用于基因作图和基因的位点定向性插入。

[0015] W000/46386 描述通过 I-SceI 双链断裂而修饰、修复、减弱和失活细胞内的基因或其它染色体 DNA 的方法。还公开了治疗或预防有需要的个体内遗传疾病的方法。还公开了嵌合的限制性内切核酸酶。

[0016] Chilton 和 Que (2003, Plant Physiol. 133 : 第 956-965 页) 和 Tzifira 等 (2003, Plant Physiol. 133 : 第 1011-1023 页) 报道 T-DNA 偏好性地整合至由切点罕见酶 I-SceI 或 I-CeuI 人工诱导的双链 DNA 断裂内。该报道还包括包含相应切点罕见酶的识别位点的供体 T-DNA 载体。

[0017] 然而,现有技术内的方法常常依赖于通过同源重组重新形成或产生完整的选择标记基因或筛选标记基因。

[0018] 因此,仍需要允许通过替换性 DNA 定向交换几乎任何靶 DNA 序列的方法。这些问题和其它问题如下文所述在本发明不同的详述实施方案内并在权利要求书内得到解决。

[0019] 发明简述

[0020] 在本发明的一个实施方案中,描述了用于将目的 DNA 分子导入植物细胞或植物的基因组内,随后去除目的 DNA 分子的亚序列(优选地包含选择标记或筛选标记)的方法,该方法包括步骤:

[0021] a. 将目的 DNA 分子导入植物细胞的基因组,目的 DNA 分子包含侧翼为同向重复排列的两个 DNA 序列的 DNA 分子的亚序列并且还包含至少一个切点罕见双链 DNA 断裂诱导 (DSBI) 酶的识别位点,其中所述识别位点位于以同向重复排列的两个 DNA 序列附近,优选地位于它们之间;

[0022] b. 选择在其中目的 DNA 分子整合进入基因组内的植物细胞并从植物细胞再生植物;

[0023] c. 将此植物与包含编码 DSBI 酶的嵌合基因的第二个植物杂交,其中所述嵌合基因包含如下有效连接的 DNA 区段:

[0024] i 小孢子特异性启动子片段,如选自核苷酸序列 SEQ ID No. 3 的启动子片段;

[0025] ii DNA 区域,其编码识别所述识别位点的切点罕见双链断裂诱导酶,如选自 I-Sce I、I-Chu I、I-Dmo I、I-Cre I、I-Csm I、PI-Fli I、Pt-Mtu I、I-Ceu I、I-Sce II、I-Sce III、H0、PI-CivI、PI-Ctr I、PI-Aae I、PI-BSU I、PI-DhaI、PI-Dra I、PI-MavI、PI-Mch I、PI-Mfu I、PI-Mfl I、PI-Mga I、PI-Mgo I、PI-MinI、PI-Mka I、PI-Mle I、PI-Mma I、PI-Msh I、PI-Msm I、PI-Mth I、PI-Mtu I、PI-Mxe I、PI-Npu I、PI-Pfu I、PI-RmaI、PI-Spb I、PI-Ssp I、PI-Fac I、PI-Mja I、PI-Pho I、PI-TagI、PI-Thy I、PI-Tko I 或 PI-Tsp I 的内切核酸酶或包含 Zn 指 DNA 结合结构域及 DNA 切割结构域的嵌合内切核酸酶;

[0026] iii 转录终止和聚腺苷酸化区域;

[0027] d. 选择包含目的 DNA 分子和编码 DSBI 酶的嵌合基因的后代植物 (F1 植物);

[0028] e. 将后代植物与另一植物杂交,由此使用后代植物作为花粉供体;

[0029] f. 选择包含编码 DSBI 酶的嵌合基因的后代植物的群体 (F2 群体);和

[0030] g. 选择在其中 DNA 分子的亚序列已经通过以同向重复排列的两个 DNA 序列间的同源重组被缺失的后代植物,并且任选地

[0031] h. 将其中 DNA 分子的亚序列已经缺失的后代植物与另一植物杂交;和

[0032] i. 得到后代植物 (F3 植物) 的群体并选择不含编码切点罕见的 DSBI 酶的嵌合基因的植物。

[0033] 在本发明的另一个实施方案中,提供用于将植物基因组、特别是植物的核基因组内的靶 DNA 序列交换为目的 DNA 序列的方法,该方法包含如下步骤:

[0034] a. 在植物细胞基因组内预先选择位点处诱导第一个双链 DNA 断裂,所述预先选择位点位于靶 DNA 序列内或位于靶 DNA 序列附近;

[0035] b. 将目的 DNA 分子导入植物细胞,该 DNA 分子包含:

[0036] i. 位于两个侧翼 DNA 区域之间的目的 DNA 序列,其中所述的侧翼 DNA 区域与植物细胞基因组内靶 DNA 序列的侧翼 DNA 区域并且优选预先选择位点的侧翼 DNA 区域具有至少 80% 序列同源性、优选 100% 序列同源性;

[0037] ii. 位于侧翼 DNA 区域之间的选择标记基因或筛选标记基因,该选择标记基因或筛选标记基因还位于侧翼 DNA 区域之一与以同向重复方式存在的部分侧翼 DNA 区域之间,其中所述的部分侧翼 DNA 区域包含侧翼 DNA 区域之一的一部分;

[0038] iii. 位于侧翼 DNA 区域之一与以同向重复方式存在的部分侧翼 DNA 区域之间的 DSBI 酶的识别位点;

[0039] c. 选择包含选择标记或筛选标记的植物细胞的群体;

[0040] d. 选择在其中目的 DNA 序列 (和选择标记或筛选标记) 已经借助侧翼 DNA 区域通过同源重组导入的植物细胞,并从植物细胞再生植物;

[0041] e. 将包含选择标记基因的再生植物或其后代植物与包含编码切点罕见双链断裂诱导 (“DSBI”) 酶的嵌合基因的植物杂交,其中该嵌合基因包含如下有效连接的 DNA 区段:

[0042] i. 小孢子特异性启动子;

[0043] ii. 编码识别位于目的 DNA 内的识别位点的双链 DNA 断裂诱导酶的 DNA 区域;

[0044] iii. 转录终止和聚腺苷酸化区域;

[0045] f. 选择包含选择标记基因或筛选标记基因和编码 DSBI 酶的嵌合基因的后代植物 (F1 植物);

[0046] g. 将后代植物与另一植物杂交,由此使用后代植物作为花粉供体;

[0047] h. 选择包含编码 DSB1 酶的嵌合基因的后代植物的群体 (F2 群体);和

[0048] i. 选择这样的后代植物,在其中选择标记基因或筛选标记基因通过侧翼 DNA 区域之一与包含侧翼 DNA 区域之一的一部分的部分侧翼 DNA 区域间的同源重组被缺失。

[0049] 本发明涉及通过以上所述方法可得到的植物。

[0050] 在本发明的又一个实施方案中,本发明涉及包含编码切点罕见的 DSB1 酶的嵌合基因,如 SEQ ID NO 6 中从核苷酸 1941 至核苷酸 3913 的嵌合基因的植物,该嵌合基因包含如下有效连接的 DNA 区段:

[0051] i. 小孢子特异性启动子,如小孢子特异性启动子片段,如选自 SEQID No 3 的核苷酸序列的启动子或其功能性片段;

[0052] ii. DNA 区域,其编码识别位于目的 DNA 内的识别位点的双链 DNA 断裂诱导酶,如选自 I-Sce I、I-Chu I、I-Dmo I、I-Cre I、I-Csm I、PI-Fli I、Pt-Mtu I、I-Ceu I、I-Sce II、I-Sce III、H0、PI-Civ I、PI-Ctr I、PI-Aae I、PI-BSU I、PI-DhaI、PI-Dra I、PI-Mav I、PI-Mch I、PI-Mfu I、PI-Mfl I、PI-Mga I、PI-Mgo I、PI-Min I、PI-Mka I、PI-Mle I、PI-Mma I、PI-Msh I、PI-MsmI、PI-Mth I、PI-Mtu I、PI-Mxe I、PI-Npu I、PI-Pfu I、PI-RmaI、PI-Spb I、PI-Ssp I、PI-Fac I、PI-Mja I、PI-Pho I、PI-TagI、PI-Thy I、PI-Tko I 或 PI-Tsp I 的内切核酸酶或包含 Zn 指 DNA 结合结构域及 DNA 切割结构域、特别是包含 SEQ ID No 1 或 SEQ ID No 2 的核苷酸序列的 DNA 区域的嵌合内切核酸酶;和

[0053] iii. 转录终止和聚腺苷酸化区域。

[0054] 本发明还涉及以上所述的嵌合基因。

[0055] 在本发明的另一个实施方案中,提供如此 DNA 载体,其借助在靶序列内或其附近预先选择位点处诱导双链断裂而将植物细胞基因组内的靶 DNA 序列交换为目的 DNA 序列,该 DNA 载体包含:

[0056] a. 位于两个侧翼 DNA 区域之间的目的 DNA 序列,其中所述的侧翼 DNA 区域与靶 DNA 序列的侧翼 DNA 区域和预先选择位点的侧翼 DNA 区域具有至少 80% 序列同源性、优选 100% 序列同源性;

[0057] b. 位于侧翼 DNA 区域之间的选择标记基因或筛选标记基因,该选择标记基因或筛选标记基因还位于侧翼 DNA 区域之一与以同向重复方式存在的包含侧翼 DNA 区域之一的一部分的部分侧翼 DNA 区域之间;和

[0058] c. 位于侧翼 DNA 区域之一与以同向重复方式存在的部分侧翼 DNA 区域之间的 DSB1 酶的识别位点。

[0059] 附图简述

[0060] 图 1 至 3 代表用于去除导入或已经导入植物细胞内的目的 DNA 中所选择亚部分的方法的不同实施方案。这些实施方案仅用于说明目的并且不应当用来以限制性方式解释权利要求书。

[0061] 图 1 是用于将具有所选择亚部分的目的 DNA 导入植物细胞并随后去除目的 DNA 中所选择亚部分的方法的示意图,其中所述的亚部分包含选择标记基因或筛选标记基因

[0062] 特征:代表任何目的 DNA 序列;DSB:双链断裂诱导酶(“DSBIE”)的识别位点;SMG1:选择标记基因或筛选标记基因;drs:同向重复序列;SMG2:与编码 DSBIE 的嵌合基因

连接的选择标记基因或筛选标记基因 ;MSP :小孢子特异性启动子 ;3' :转录终止和聚腺苷酸化信号 ;

[0063] 图 2 是允许用替换性 DNA 序列精确替换靶 DNA 序列的方法的示意图。DSB1 :第一个双链断裂诱导酶的识别位点 ;FS1 :侧翼序列 1 ;FS2 :侧翼序列 2 ;DSB2 :第二个双链断裂诱导酶的识别位点 ;SMG1 :选择标记基因 1 或筛选标记基因 1 ;SMG2 :选择标记基因 2 或筛选标记基因 2 ;DSBIE :双链断裂诱导酶 ;dr1 :同向重复序列 1 (其与作为侧翼序列 2 的一部分的同向重复序列 2 相似或相同 ;本文中称作“部分侧翼 DNA 区域”) ;MSP :小孢子特异性启动子 ;3' :转录终止和聚腺苷酸化信号。

[0064] 图 3 是与图 2 内所示方法类似的允许用替换性 DNA 序列精确替换靶 DNA 序列的方法的示意图。dr1 在此情况下是这样的同向重复序列,其是来自侧翼序列 1 的一部分并与同向重复序列 2(dr2) 相似或相同。

[0065] 发明的详细实施方案

[0066] 本发明基于如此发现,即可以高效地去除侧翼为两个同向重复序列并位于切点罕见双链 DNA 断裂诱导酶的识别位点附近的 DNA 分子中所选择序列,这在当包含此 DNA 的植物首先与包含处于小孢子特异性启动子控制下的编码切点罕见双链 DNA 断裂诱导酶的嵌合基因的植物杂交并将所得植物的花药用来自受体植物授粉时实现。

[0067] 因此,本发明在一个实施方案内涉及包含处于小孢子特异性启动子的控制下的、编码切点罕见双链 DNA 断裂诱导内切核酸酶的嵌合基因的植物用途,以便通过杂交去除位于切点罕见双链 DNA 断裂诱导内切核酸酶的识别位点附近并还位于以同向重复方向存在的两个序列之间的 DNA 片段 (见图 1)。切点罕见的 DSBIE 内切核酸酶在花药形成期间于小孢子内的表达足以诱导双链 DNA 断裂并因而显著刺激同向重复序列间的染色体同源重组,导致去除位于这些同向重复序列之间的序列。

[0068] 也就是说,在本发明的一个实施方案中,提供用于将目的 DNA 分子导入植物细胞或植物的基因组内,随后去除该 DNA 分子的亚序列的方法,该方法包括步骤:

[0069] a. 将此目的 DNA 分子导入植物细胞的基因组,此目的 DNA 分子包含侧翼为同向重复排列的两个 DNA 序列的该 DNA 分子的亚序列,并且还包含至少一个切点罕见双链 DNA 断裂诱导 (DSBIE) 酶的识别位点,其中所述的识别位点位于以同向重复排列的两个 DNA 序列之间 ;

[0070] b. 选择在其中目的 DNA 分子整合进入基因组内的植物细胞并从植物细胞再生植物 ;

[0071] c. 将此植物与包含编码 DSBIE 酶的嵌合基因的第二个植物杂交,其中所述的嵌合基因包含如下有效连接的 DNA 区段 :

[0072] i. 小孢子特异性启动子 ;

[0073] ii. 编码识别所述识别位点的切点罕见双链断裂诱导酶的 DNA 区域 ;

[0074] iii. 转录终止和聚腺苷酸化区域 ;

[0075] d. 选择包含目的 DNA 分子和编码 DSBIE 酶的嵌合基因的后代植物 (F1 植物) ;

[0076] e. 将后代植物与另一植物杂交,由此使用后代植物作为花粉供体 ;

[0077] f. 选择包含编码 DSBIE 酶的嵌合基因的后代植物的群体 (F2 群体) ;和

[0078] g. 选择在其中目的 DNA 分子的亚序列已经通过以同向重复排列的两个 DNA 序列间

的同源重组被缺失的后代植物。

[0079] 如本文中所用,“切点罕见双链 DNA 断裂诱导内切核酸酶”是能够在称作“识别位点”的特定核苷酸序列内诱导双链 DNA 断裂的酶。切点罕见内切核酸酶,有时又称作大范围核酸酶,具有 14 至 40 个连续核苷酸的识别位点。因此,切点罕见内切核酸酶具有非常低的切割频率,即便在较大的植物基因组内亦如此。归巢内切核酸酶构成此类切点罕见内切核酸酶的一个家族。它们可以由内含子、独立基因或间插序列编码,并展示出使之与通常来自细菌限制性修饰 II 型系统更常见的限制酶相区别的醒目结构特性和功能特性。它们的识别位点具有广泛非对称性,这与绝大多数限制酶识别位点的特征性二重对称相反。已经证实由内含子或内含肽编码的数种归巢内切核酸酶可促进它们各自的遗传元件归巢至无内含子或无内含肽的等位位点内。通过无内含子或无内含肽的等位基因内产生位点特异性双链断裂,这些核酸酶产生重组性末端,该重组末端参与了复制编码序列并在 DNA 水平引起内含子或间插序列插入的基因转变过程。

[0080] 充分得到表征的归巢内切核酸酶是 I-SceI。I-SceI 是酿酒酵母 (*Saccharomyces cerevisiae*) 内负责线粒体中内含子移动的位点特异性内切核酸酶。该酶由 21S rRNA 基因的任选内含子 Sc LSU.1 编码并在内含子插入位点处诱导双链 DNA 断裂,产生带 3'OH 突出端的 4bp 交错切口。I-SceI 内切核酸酶的识别位点延伸超过 18bp 非对称性序列 (Colleaux 等 1988 Proc. Natl. Acad. Sci. USA 85:6022-6026)。I-SceI 的氨基酸序列和与线粒体 I-SceI 基因等效的通用密码子已经例如由 WO 96/14408 提供。WO96/14408 也公开了仍有功能的 I-SceI 蛋白的众多变体。

[0081] PCT 申请 PCT/EP04/013122 (本文中引用作为参考) 提供已经优化用于植物中表达的 I-SceI 的合成性核苷酸序列变体。此合成性 I-SceI 编码区的核苷酸序列以 UIPAC 密码子描述于 SEQ ID No 1。UIPAC 密码子的符号具有它们的常用含义,即 N = A 或 C 或 G 或 T; R = A 或 G; Y = C 或 T; B = C 或 G 或 T (非 A); V = A 或 C 或 G (非 T); D = A 或 G 或 T (非 C); H = A 或 C 或 T (非 G); K = G 或 T; M = A 或 C; S = G 或 C; W = A 或 T。

[0082] 其它切点罕见的 DSB 酶及它们相应的识别位点的列表提供于 W003/004659 的表 1 内 (第 17 至第 20 页) (本文中引用作为参考)。它们包括 I-SceI、I-Chu I、I-Dmo I、I-Cre I、I-Csm I、PI-Fli I、Pt-Mtu I、I-Ceu I、I-SceII、I-Sce III、H0、PI-Civ I、PI-Ctr I、PI-Aae I、PI-BSU I、PI-DhaI、PI-Dra I、PI-Mav I、PI-Mch I、PI-Mfu I、PI-Mfl I、PI-Mga I、PI-Mgo I、PI-Min I、PI-Mka I、PI-Mle I、PI-Mma I、PI-Msh I、PI-Msm I、PI-MthI、PI-Mtu I、PI-Mxe I、PI-Npu I、PI-Pfu I、PI-Rma I、PI-Spb I、PI-SspI、PI-Fac I、PI-Mja I、PI-Pho I、PI-Tag I、PI-Thy I、PI-Tko I 或 PI-TspI。

[0083] 此外,可获得设计基本上识别任何所选择靶核苷酸序列的客户定制型切点罕见内切核酸酶的方法。简而言之,嵌合性限制酶可以使用为识别特定核苷酸序列所设计的锌指结构域与来自天然限制酶如 FokI 的非特异性 DNA 切割结构域的杂合体加以制备。此类方法已经描述于 WO 03/080809、WO94/18313 或 WO95/09233 以及 Isalan 等, 2001, Nature Biotechnology 19,656-660; Liu 等 1997, Proc. Natl. Acad. Sci. USA 94, 5525-5530 内。通过从变体的文库中选择而产生客户定制型大范围核酸酶的另一个途径描述于 W02004/067736 内。

[0084] 如本文中所用,“侧翼为同向重复排列的两个 DNA 序列”表示两个 DNA 区域分别紧

接在待从导入的 DNA 分子中去除的序列的每一端,其中所述的两个 DNA 区域在核苷酸序列上基本上相似。同向重复序列不必完全相同,但是可以在约 75%至约 100%序列同一性之间变化。重复序列越短,对序列相似性的要求越严格。然而,为了恢复 DNA 序列而不留下痕迹,如下文中所述,以同向重复排列的 DNA 序列优选应当完全相同。为避免疑问,如果在核苷酸序列上基本上相似的两个 DNA 区域包含于双链 DNA 分子内,则这些 DNA 序列应当位于同一 DNA 链上,处于相同的 5' → 3' 方向。

[0085] 重复的 DNA 序列可以是至少 10 个、50 个或 100 个核苷酸长度,不过,序列当然可以更长。然而,已经发现长度超过 300 个核苷酸的重复序列不再显著地增强导致位于同向重复序列之间 DNA 序列去除的染色体同源重组。

[0086] 为本发明的目的,表述为百分数的两个相关性核苷酸序列或氨基酸序列的“序列同一性”指在两个优化比对序列内具有完全相同残基的位置数除以所比较的位置数($\times 100$)。将空位(即对比中一个残基存在于一个序列内而不存在于另一序列的位置)视作不具有完全相同残基的位置。两个序列的对比通过 Needleman 和 Wunsch 算法(Needleman 和 Wunsch 1970)开展。计算机辅助的序列对比可以使用标准软件程序,如作为 Wisconsin Package Version 10.1(Genetics Computer Group, Madison, Wisconsin, USA) 部分的 GAP, 利用具有空位生成罚值 50 和空位延伸罚值 3 的默认计分矩阵方便地开展。

[0087] 虽然 DSB1 识别位点优选地位于同向重复的 DNA 序列之间,但是这既非重要,也不需要。实际上,DSB1 识别位点还可能是重复 DNA 序列中一个序列的一部分。

[0088] 如本文中所用“位于附近”指 DSB1 的位置距离同向重复 DNA 序列在 500bp、1kbp 至 10kbp 之间。

[0089] 本文中所述的方法需要利用编码切点罕见双链断裂诱导酶的嵌合基因,由此内切核酸酶的编码区处在小孢子特异性启动子片段的控制下。

[0090] 如本文中所用的“小孢子特异性启动子区域”或“小孢子特异性启动子”或“小孢子特异性启动子片段”是能够选择性地、优选是特异性地在植物的单细胞性小孢子内促进转录的启动子区域或启动子或启动子片段。在被子植物中,有性繁殖需要产生有活力的雄性配子体和雌性配子体。作为雄性配子体的花粉在花药内形成并从造孢细胞开始,发育成性母细胞。性母细胞发生减数分裂以形成单倍体细胞的四分体,四分体随后释放至花药室。随着扩张和空泡形成,小孢子的非对称性有丝分裂产生含有营养细胞和生殖细胞的双细胞性花粉。在大部分物种内,花药在双细胞条件下散发。在 WO 97/30166(本文中引用作为参考;还见 SEQ ID No 3)内描述了作为合适小孢子特异性启动子区域的烟草 NTM19 基因启动子区域。其功能性片段已经整合入实施例的嵌合基因(SEQ ID No 6)内。小孢子特异性启动子片段可以包括 SEQ ID No 3 中从位置 1 至位置 954 或从位置 1 至位置 993 的核苷酸序列或 SEQ ID No 6 中从位置 1941 至 2926 的核苷酸序列。

[0091] 如本文中所用“切点罕见双链断裂诱导内切核酸酶的编码区”或“切点罕见双链断裂诱导酶的编码区”是编码特征为切点罕见的 DSB1 酶,如归巢内切核酸酶或在本申请内其它处所描述的嵌合内切核酸酶的核苷酸序列。编码区因此可以包含编码下表内所列归巢内切核酸酶中任意氨基酸序列的任意核苷酸序列,其中所述的归巢内切核酸酶可以在提到的登录号下于公共数据库内找到(它们在本文中均引用作为参考):

[0092]

DSBI 酶	登录号
I-AniI	P03880
I-CvuI	P56347
I-CreI	P05725
I-ChuI	Q32001
I-CpaI-I-CpaIII-I-CpaIV-I-CpaV	Q39562/Q8WKZ5/Q8WKZ6/Q8WKZ8

[0093]

I-CpaII	Q39559
I-CeuI	P32761
I-DmoI	P21505
I-SceI	P03882
I-SceII	P03878
I-SceIII	Q9ZZX3
PI-SceI	P17255
I-NanI	Q25535
I-NitI	Q25567
I-NjaI	Q25568
I-PpoI	Q94702
PI-PfuI	073954
PI-PkoI	P77933
PI-PkoII	P77933
PI-PspI	Q51334
PI-TfuI	P74918
PI-TfuII	P74918

PI-ThyI	Q9HH05
PI-ThyII	Q9HH05
PI-TliI	P30317
PI-TliII	P30317
I-TevI	P13299
I-TevII	P07072
I-TevIII	Q38419

[0094] 显而易见,为了在小孢子特异性启动子片段的控制下表达内切核酸酶,应当调整编码区以便使用通用密码子语言编码以上提到的多肽。编码区还可以优化用于在植物内表达,并且合成性编码区具有为满足如下标准而设计的核苷酸序列:

[0095] a) 核苷酸序列编码有功能的切点罕见双链断裂诱导内切核酸酶,

[0096] b) 核苷酸序列具有约 50%至约 60%的 GC 含量,

[0097] c) 核苷酸序列不包含选自 GATAAT、TATAAA、AATATA、AATATT、GATAAA、AATGAA、AATAAG、AATAAA、AATAAT、AACCAA、ATATAA、AATCAA、ATACTA、ATAAAA、ATGAAA、AAGCAT、ATTAAT、ATACAT、AAAATA、ATTAAA、AATTAA、AATACA 和 CATAAA 的核苷酸序列;

[0098] d) 核苷酸不包含选自 CCAAT、ATTGG、GCAAT 和 ATTGC 的核苷酸序列;

[0099] e) 核苷酸序列不包含选自 ATTTA、AAGGT、AGGTA、GGTA 或 GCAGG 的序列;

[0100] f) 核苷酸序列不包含由选自 G 或 C 的 7 个连续核苷酸组成的 GC 段;

[0101] g) 核苷酸序列不包含由选自 A 或 T 的 5 个连续核苷酸组成的 GC 段;和

[0102] h) 核苷酸序列不包含如此密码子,其编码 Leu、Ile、Val、Ser、Pro、Thr、Ala 并在位置 2 和 3 内包含 TA 或 CG 双体(即核苷酸序列不包含密码子 TTA、CTA、ATA、GTA、TCG、CCG、ACG 和 GCG)。

[0103] 双链断裂诱导酶可以包含,但不是必须包含,核定位信号(NLS)[Raikhel, Plant Physiol. 100:1627-1632(1992) 及其中的参考文献],如 SV40 大 T 抗原的 NLS[Kalderon 等 Cell 39:499-509(1984)]。核定位信号可以位于蛋白质内任何地方,但是合适地位于蛋白质的氨基末端。核定位信号可以替换双链断裂诱导酶中的一个或多个氨基酸。

[0104] 虽然用于去除的方法在本文中描述为涉及导入目的 DNA 分子的活化步骤随后去除所选择亚片段,显而易见的是本发明的去除方法可以用来去除位于 DNA 同向重复序列之间的任何序列,只要可以找到或设计出识别在重复 DNA 序列附近的 DSB1 识别位点的 DSB1 酶。

[0105] 还将显而易见的是,用来描述本方法的术语如“DNA 片段的导入”及“从细胞再生植物”不表示该 DNA 片段必定需要通过转化技术而导入。实际上对于本领域技术人员而言立刻明白目的 DNA 分子还可以通过育种或杂交技术从一种植物导入另一种植物内。

[0106] 然而,将显而易见的是目的 DNA 分子可以通过本领域已知的任何方法导入植物细胞,包括农杆菌介导性转化和 DNA 直接转移方法。可以使用任何便利的方法将转化性 DNA

分子转移至植物细胞内,包括但不限于 DNA 直接转移。如本文中所用“DNA 直接转移”是将 DNA 导入植物细胞的任何方法,该方法不涉及利用天然的农杆菌物种并能够将 DNA 导入植物细胞。这包括本领域众所周知的方法,如通过电穿孔法将 DNA 导入原生质体、通过电穿孔法将 DNA 导入完整的植物细胞或经部分降解的组织或植物细胞、通过试剂如 PEG 等的作用将 DNA 导入原生质体、利用硅晶须以及用包被 DNA 微粒轰击。

[0107] DNA 可以通过如在例如 PCT/EP04/013122 内所述的涉及在预先选择位点处导入双链断裂的同源重组或非同源末端连接的方法整合。

[0108] 在本发明的一个特定实施方案中,去除的方法可以与 DNA 通过定向同源重组而插入、缺失或替换组合地使用,并且在该方法中如此实现定向 DNA 插入,即使用选择标记或筛选标记,随后在包含选择标记或筛选标记的植物细胞或植物的群体内验证在其中定向 DNA 插入通过同源重组而发生的植物细胞或植物。当侧翼序列和同向重复序列合适地进行选择时,该方法引起靶 DNA 精确替换成目的 DNA,而不存在用来实现替换的目的 DNA 分子的任何残留(“痕迹”)。去除的方法也不需要任何额外的体外培养,因而避免产生体细胞性变异。本方法的示意图提纲可以在图 2 和 3 内找到。

[0109] 有趣的是,已经观察到使用如 PCT/EP04/013122 内所述的借助同源重组定向插入外源目的 DNA 的方法,在其中外源 DNA 确实经同源重组插入的那些转化事件代表了在其中 DNA 通过任何手段整合入植物染色体内的事件总群体中的相对高比例(在 1%至 5%的数量级上)。因此,不需要依赖产生可识别表型的 DNA 序列(如同源重组后产生的完整选择标记基因)经同源重组的生成或再造,以鉴定 DNA 因而通过同源重组得以插入的那些事件。此外,选择标记基因或筛选标记基因可以包含于侧翼 DNA 序列之间的 DNA 区域内,随后分析数目相对较少的转化的植物细胞或植物,以鉴定在其中定向 DNA 插入经同源重组而发生的那些转化事件。

[0110] 因此,在本发明的这个实施方案中,提供用于将植物细胞内的靶 DNA 序列交换成目的 DNA 序列(或外源 DNA)的方法,该方法包含如下步骤:

[0111] - 在细胞基因组内预先选择位点处诱导第一个双链 DNA 断裂,所述预先选择位点位于靶 DNA 序列内或位于该靶 DNA 序列的附近;

[0112] - 将目的 DNA 分子(或外源 DNA)导入植物细胞,由此 DNA 分子包含如下有效连接的 DNA 片段:

[0113] i. 位于两个侧翼 DNA 区域之间的目的 DNA 序列,其中所述的侧翼 DNA 区域与靶 DNA 序列的侧翼 DNA 区域和预先选择位点的侧翼 DNA 区域具有至少 80%序列同源性、优选 100%序列同源性;

[0114] ii. 位于侧翼 DNA 区域之间的选择标记基因或筛选标记基因,因而选择标记基因或筛选标记基因还位于侧翼 DNA 区域之一和以同向重复方式存在的所提及侧翼 DNA 区域之一的至少部分的另一个拷贝(又标示为部分侧翼序列)之间;

[0115] iii. 位于侧翼 DNA 区域之一与以同向重复方式存在的部分侧翼 DNA 区域之间的 DSB1 酶的识别位点;

[0116] - 选择包含选择标记或筛选标记的植物细胞的群体;

[0117] - 选择在其中选择标记或筛选标记已经借助侧翼 DNA 区域通过同源重组导入的植物细胞,并从植物细胞再生植物;

[0118] - 将包含选择标记基因的再生植物或其后代植物与包含编码 DSBI 酶的嵌合基因的植物杂交,其中该嵌合基因包含如下有效连接的 DNA 区段:

[0119] ■ 小孢子特异性启动子;

[0120] ■ 编码识别位于目的 DNA 内的识别位点的双链 DNA 断裂诱导酶的 DNA 区域;

[0121] ■ 转录终止和聚腺苷酸化区域;

[0122] - 选择包含选择标记基因或筛选标记基因和编码 DSBI 酶的嵌合基因的后代植物 (F1 植物);

[0123] - 将后代植物与另一植物杂交,由此使用后代植物作为花粉供体;

[0124] - 选择包含编码 DSBI 酶的嵌合基因的后代植物的群体 (F2 群体);

[0125] 和

[0126] - 选择所述 F2 群体内的后代植物,在所述后代植物中选择标记基因或筛选标记基因通过侧翼 DNA 区域之一与包含侧翼 DNA 区域之

[0127] 一的一部分的部分侧翼 DNA 区域间的同源重组被缺失。

[0128] 因此,如本文中所用“预先选择位点”指示植物的核基因组内位于靶 DNA 序列内或附近的特定核苷酸序列,在该位置内需要插入外源 DNA 或将靶 DNA 序列交换。本领域技术人员将能够选择识别所选择靶核苷酸序列的双链 DNA 断裂诱导 (“DSBI”) 酶或设计此种 DSBI 内切核酸酶。或者,DSBI 内切核酸酶识别位点可以使用任何常用的转化方法导入或使用在其基因组内具有 DSBI 内切核酸酶识别位的植物系通过常规育种法导入,并且任何所需要的外源 DNA 可以随后导入事先已导入的预先选则靶位点内。

[0129] 在转化性 DNA 分子内的双链 DNA 断裂可以通过植物可表达的嵌合基因的瞬时导入而方便地诱导,其中所述的嵌合基因包含与编码双链断裂诱导酶的 DNA 区域有效连接的植物可表达启动子区域。编码双链断裂诱导酶的 DNA 区域可以是合成性 DNA 区域,如 (但不限于) 这样的合成性 DNA 区域,其中密码子根据本申请内其它处所述的用于 I-SceI 编码区设计方案而选择。作为蛋白质,内切核酸酶本身还可以例如通过电穿孔法导入植物细胞。然而,内切核酸酶还可以通过将嵌合基因 (包含与诱导型植物可表达启动子有效连接的编码内切核酸酶的 DNA 区域) 导入植物细胞或植物的基因组,并在导入转化性 DNA 分子之前、期间或之后立即提供合适的诱导化合物持续所限时间而以瞬时方式提供。内切核酸酶还可以作为编码内切核酸酶的 RNA 前体提供。

[0130] 双链断裂诱导酶可以包含,但不必须包含,核定位信号 (NLS) [Raikhel, Plant Physiol. 100:1627-1632 (1992) 及其中参考文献],如 SV40 大 T 抗原的 NLS [Kalderson 等 Cell 39:499-509 (1984)]。核定位信号可以位于蛋白质内任何地方,但是合适地位于蛋白质的氨基末端。核定位信号可以替换双链断裂诱导酶中的一个或多个氨基酸。

[0131] 如本文中所用,“靶 DNA 序列”是通过添加、缺失或置换而修饰的位于植物细胞基因组内的 DNA 序列。

[0132] 如本文中所用“侧翼 DNA 区域”是分别与靶 DNA 序列上游或下游 DNA 区域具有同源性的 DNA 序列。这允许更好地控制外源 DNA 或目的 DNA 分子的插入。实际上,通过同源重组的整合将允许外源 DNA 片段与植物的核基因组在核苷酸水平上精确连接。

[0133] 侧翼 DNA 区域可以在长度上变化,并且应当是至少约 10 个核苷酸长度。然而,侧翼区域可以按照实际可能尽量地长 (例如直至约 100-150kb,如完整的细菌人工染色体

(BAC))。优选地,侧翼区域会是约 50bp 至约 2000bp。此外,侧翼为外源目的 DNA 的区域不需要与预先选择位点的侧翼 DNA 区域完全相同,并且可以与预先选择位点的侧翼 DNA 区域具有约 80%至约 100%之间的序列同一性,优选地约 95%至约 100%序列同一性。侧翼区域越长,对同源性要求的严格性越低。此外,优选序列同一性在靠近精确插入外源 DNA 的位置处按照实际可能尽量地高。此外,为实现交换靶 DNA 序列而不改变毗邻 DNA 序列的 DNA 序列,侧翼 DNA 序列应当优选地与预先选择位点侧翼的 DNA 区域完全相同。

[0134] 此外,外源目的 DNA 侧翼的区域不必与紧邻预先选择位点侧翼的区域具有同源性,但是可以与距离预先选择位点更远的核基因组的 DNA 区域具有同源性。外源 DNA 的插入随后将导致在预先所选择插入位点和同源 DNA 区域之间的靶 DNA 的去除。也就是说,位于同源区域之间的靶 DNA 将置换成外源目的 DNA。

[0135] 优选地,预先选择位点和另外提及的识别序列由不同的切点罕见双链断裂诱导内切核酸酶识别。

[0136] 提及的“部分侧翼 DNA 区域”指邻近待缺失且通常包含选择标记或筛选标记的 DNA 区域的侧翼 DNA 区域中至少部分的 DNA 区域。很明显,部分侧翼 DNA 序列也可以在长度上与侧翼 DNA 序列相等或甚至包含更长的侧翼 DNA 序列。

[0137] “选择标记或筛选标记”如本文中所用具有本领域的通常含义并包括,但不限于植物可表达的蒴丝菌素乙酰转移酶、新霉素磷酸转移酶、草甘膦氧化酶、草甘膦耐受 EPSP 酶、腈水解酶基因、突变的乙酰乳酸合酶或乙酰羟酸合酶基因、 β -葡萄糖醛酸糖苷酶 (GUS)、R- 基因座基因、绿色荧光蛋白等。

[0138] 选择在其中选择标记或筛选标记和外源 DNA 分子的剩余部分已经借助侧翼 DNA 区域通过同源重组得以导入的植物细胞或植物例如可以通过对如此序列的缺乏进行筛选而实现,其中所述序列存在于转化性 DNA 内,但位于侧翼 DNA 区域之外。实际上,来自转化性 DNA 但位于侧翼 DNA 区域之外的序列的存在表示着为来源于 DNA 随机插入的转化植物细胞。为此目的,选择标记或筛选标记可以包含于侧翼 DNA 区域之外的转化性 DNA 分子内,转化性 DNA 分子随后可以用于鉴定不具有位于转化性 DNA 之外的并且已经借助侧翼 DNA 区域通过同源重组而产生的选择标记或筛选标记的植物细胞。备选地,转化性 DNA 分子可以含有允许对缺少此类基因(负选择标记基因)进行选择、位于侧翼 DNA 区域之外的选择标记。

[0139] 在本发明的另一个实施方案中,本文中所述的 DNA 去除方法可以与如此方法组合,即该方法以非同源末端连接为基础用于在细胞基因组内预先选择位点处插入 DNA。

[0140] 因此,本发明提供用于在基因组内、优选是植物细胞的核基因组内的预定位置处插入所选择 DNA 分子的方法,该方法包括如下步骤:

[0141] - 在细胞基因组内预先选择位点处诱导第一个双链 DNA 断裂,所述预先选择位点优选地位于靶 DNA 序列内;

[0142] - 将外源 DNA 分子导入植物细胞,因而该 DNA 分子包含如下有效连接的 DNA 片段:

[0143] ○所选择目的 DNA 分子;

[0144] ○在前面或后面具有重复 DNA 区域的选择标记基因或筛选标记基因,其中所述的重复 DNA 区域与位置紧邻预先选择位点的基因组 DNA 区域之一具有至少 80%的序列同一性,因而当外源 DNA 分子通过非同源末端连接在预先选择位点内插入时,该 DNA 区域与其基因组拷贝以同向重复方式存在;

[0145] ○切点罕见的 DSBI 酶的识别位点,位于包含所述重复 DNA 区域和所述选择标记基因的外源 DNA 区域内;

[0146] - 选择包含选择标记或筛选标记的植物细胞的群体;

[0147] - 选择在其中选择标记或筛选标记已经通过非同源末端连接在预先选择位点处导入的植物细胞,并从植物细胞再生植物;

[0148] - 将包含选择标记基因的再生植物或其后代植物与包含编码 DSBI 酶的嵌合基因的植物杂交,其中该嵌合基因包含如下有效连接的 DNA 区段:

[0149] ■小孢子特异性启动子;

[0150] ■编码识别位于目的 DNA 内的识别位点的双链 DNA 断裂诱导酶的 DNA 区域;

[0151] ■转录终止和聚腺苷酸化区域;

[0152] - 选择包含选择标记基因或筛选标记基因和编码 DSBI 酶的嵌合基因的后代植物(F1 植物);

[0153] - 将后代植物与另一植物杂交,由此使用后代植物作为花粉供体;

[0154] - 选择包含编码 DSBI 酶的嵌合基因的后代植物的群体(F2 群体);和

[0155] - 选择所述 F2 群体内的后代植物,在所述后代植物中选择标记基因或筛选标记基因通过重复 DNA 区域和位置紧邻预先选择位点的基因组 DNA 区域间的同源重组被缺失。

[0156] 以上提及的方法可以方便地用于打断所选择的任何 DNA 序列,例如多肽编码区、编码生物活性 RNA 的 DNA 序列、启动子区域、调节区、蛋白质或 RNA 结合的识别位点等。

[0157] 在该实施方案中,在其中 DNA 分子已经通过非同源末端连接而插入的事件可以例如使用引物序列(识别位于预先选择位点附近的基因组序列并且还优选地不识别识别外源 DNA)和在外源 DNA 分子内的引物通过 PCR 反应方便地进行鉴定。在通过非同源末端连接在预先选择位点处插入外源 DNA 时,DNA 片段将扩增。当外源 DNA 随机整合时,不会扩增此 DNA 片段。

[0158] 可以理解的是本发明的手段和方法可以用于能够通过花粉繁殖的任何植物内,包括玉米、烟草、谷物植物包括小麦、燕麦、大麦、黑麦、稻、草坪草、高粱、谷子或甘蔗。本发明的方法还可以应用于任何植物(被子植物或裸子植物),包括但不限于棉花、油菜、欧洲油菜、大豆、蔬菜、马铃薯、浮萍属某些种(*Lemna* spp.)、烟草属某些种(*Nicotiana* spp.)、拟南芥(*Arabidopsis*)、紫花苜蓿、大麦、菜豆、玉米、棉花、亚麻、豌豆、芸苔、稻、黑麦、红花、高粱、大豆、向日葵、烟草、小麦、芦笋、甜菜、西兰花、甘蓝、胡萝卜、花椰菜、芹菜、黄瓜、茄子、莴苣、洋葱、欧洲油菜、胡椒、马铃薯、西葫芦、萝卜、菠菜、南瓜、番茄、小胡瓜、扁桃、苹果、杏、香蕉、黑莓、蓝莓、可可、樱桃、椰子、蔓越橘、枣、葡萄、葡萄柚、番石榴、猕猴桃、柠檬、酸橙、芒果、甜瓜、油桃、桔、番木瓜、西番莲、桃、花生、梨、菠萝、阿月浑子、李、树莓、草莓、橘、胡桃和西瓜。

[0159] 本发明的目的还是提供根据本发明方法所产生的植物细胞和植物。本发明的范围还包括通过传统育种方法产生的包含 DNA 插入事件的植物的配子、种子、胚、合子性或体细胞性后代或杂种。此类植物可以含有替代靶序列的异源 DNA 序列并且仅仅因交换后这种异源 DNA 或 DNA 序列的存在而与它们的祖先植物不同。

[0160] 通过本文中所述方法得到的植物可以通过常规育种技术与其它植物进一步杂交,以得到包含根据本发明所得到的定向 DNA 插入事件的后代植物。

[0161] 如下非限制性实施例描述使用处于小孢子特异性启动子控制下表达的编码双链 DNA 断裂诱导酶如 I-SceI 的嵌合基因从所导入 DNA 分子中去除所选择亚片段。

[0162] 在实施例中除非另外说明,所有重组 DNA 技术根据如在 Sambrook 等 (1989) Molecular Cloning :A Laboratory Manual, 第二版, Cold Spring Harbor Laboratory Press, NY 和在 Ausubel 等 (1994) Current Protocols in Molecular Biology, Current Protocols, USA 的第 1 和 2 卷内所述的标准方案开展。用于植物分子工作的标准材料和方法描述于 BIOS Scientific Publications Ltd(UK) 和 Blackwell Scientific Publications, UK 联合出版的 R. D. D. Croy 编的 Plant Molecular Biology Labfax(1993) 内。对于标准分子生物学技术的其它参考文献包括 Sambrook 和 Russell (2001) Molecular Cloning :A Laboratory Manual, 第三版, Cold Spring Harbor Laboratory Press, NY、Brown(1998) Molecular Biology LabFax, 第二版, Academic Press(UK) 第 I 和 II 卷。用于聚合酶链式反应的标准材料和方法可以在 Dieffenbach 和 Dveksler (1995) PCR Primer : A Laboratory Manual, Cold Spring Harbor Laboratory Press 并在 McPherson 等 (2000) PCR-Basics :From Background to Bench, 第一版, Springer Verlag, Germany 内找到。

[0163] 在说明书和实施例通篇范围内,提到如下序列:

[0164] SEQ ID No 1 :合成性 I-SceI 编码区 (UIPAC 编码) 的核苷酸序列。

[0165] SEQ ID No 2 :合成性 I-SceI 编码区的核苷酸序列。

[0166] SQ ID No 3 :包括启动子区域的小孢子选择性 NTM19 基因的核苷酸序列。

[0167] SEQ ID No 4 :pTCV63 的 T-DNA 的核苷酸序列。

[0168] SEQ ID No 5 :pTCV64 的 T-DNA 的核苷酸序列。

[0169] SEQ ID No 6 :pTCV72 的 T-DNA 的核苷酸序列。

实施例

[0170] 通过染色体内同源重组 (IHR) 去除选择标记基因

[0171] 已经开发出检测去除所选择 DNA 片段的重组分析法,其基于在同向重复序列 (egfp 序列的一部分 ;约 300bp 或约 600bp) 间的染色体内同源重组 (IHR) 而去除选择标记基因 (hyg) (~ 2000bp) 后恢复 egfp-bar 融合基因。重复序列之一的侧翼具有 I-SceI (和锌指 Zif268) 识别位点,这有可能在重复序列间产生 DSB。为了在一个世代转移至另一个世代期间允许 IHR,将 I-SceI 内切核酸酶置于小孢子特异性启动子的控制下 (pNTM19)。

[0172] 使用标准重组 DNA 技术,构建以下 DNA 分子用于如下的实验:

[0173] 1. pTCV63 :具有含如下有效连接的 DNA 构建体的同向重复短序列 (~ 300bp) :

[0174] -p35S3 :CaMV35S 启动子片段

[0175] -egf (短) :eGFP 编码序列的第一个部分,包含与随后名为 GFP 序列重叠的 300bp

[0176] -I-SceI 内切核酸酶的识别位点

[0177] -含 Zn 指 Zif268 的 DNA 结合蛋白的识别位点

[0178] -pCsVMV :木薯叶脉花叶病毒启动子片段

[0179] -hyg :潮霉素抗性编码区

[0180] -3' 35S :3' 转录终止和聚腺苷酸化信号

[0181] -gfp (短) :eGFP 编码序列的 3' 部分,包含该质粒的先前 egf 部分的 300bp 的同

向重复序列,并且在其中编码区翻译性地与 bar 基因编码区连接。

[0182] -3' nos :来自胭脂氨酸合酶基因的 3' 转录终止和聚腺苷酸化信号。

[0183] 将该质粒导入根瘤农杆菌 (*Agrobacterium tumefaciens*) 并且使用得到的菌株 (A4330) 来产生转基因烟草植物 (G7NT001)。

[0184] 2. pTCV64 :具有含如下有效连接的 DNA 构建体的同向重复长序列 (~ 600bp) :

[0185] -p35S3 :CaMV35S 启动子

[0186] -egf (长) :eGFP 编码序列的第一个部分,包含与随后名为 gfp 序列重叠的 600bp

[0187] -I-SceI 内切核酸酶的识别位点

[0188] -含 Zn 指 Zif268 的 DNA 结合蛋白的识别位点

[0189] -pCsVMV :木薯叶脉花叶病毒启动子

[0190] -hyg :潮霉素抗性编码区

[0191] -3' 35S :3' 转录终止和聚腺苷酸化信号

[0192] -gfp (长) :efgp 编码序列的 3' 部分,包含先前 egf 构建体的 600bp 的同向重复序列,并且在其中编码区翻译性地与 bar 基因编码区连接。

[0193] -3' nos :来自胭脂氨酸合酶基因的 3' 转录终止和聚腺苷酸化信号。

[0194] 将该质粒导入根瘤农杆菌并且使用得到的菌株 (A4364) 来产生转基因烟草植物 (G7NT004)。

[0195] 3. pTCV72 :

[0196] -pnos :胭脂氨酸合酶启动子

[0197] -neo :新霉素磷酸转移酶 II 编码区

[0198] -3' ocs :来自章鱼碱合酶基因的 3' 转录终止和聚腺苷酸化信号 ;

[0199] -pNTM19 :小孢子特异性启动子片段

[0200] -I-SceI :内切核酸酶 I-SceI 的编码区

[0201] -3' nos :来自 CaMV 35S 转录物的 3' 转录终止和聚腺苷酸化信号。

[0202] 将该质粒导入根瘤农杆菌并且使用得到的菌株 (A4331) 来产生转基因烟草植物 (G7NT005)。

[0203] 从 G7NT001 和 G7NT004 各自的三个独立的单拷贝转化烟草系中,使用 G7NT005 作为雄性植物开展与包含编码 I-SceI 的嵌合基因的两个独立的单拷贝转化系 (G7NT005) 杂交,其中所述的嵌合基因处于小孢子特异性启动子控制下,因而后代系如下所示 :

[0204] G7NT001-0001 × G7NT005-0001 > 04TDNT000001

[0205] G7NT001-0002 × G7NT005-0001 > 04TDNT000002

[0206] G7NT001-0003 × G7NT005-0001 > 04TDNT000003

[0207] G7NT001-0001 × G7NT005-0002 > 04TDNT000004

[0208] G7NT001-0002 × G7NT005-0002 > 04TDNT000005

[0209] G7NT001-0003 × G7NT005-0002 > 04TDNT000006

[0210] G7NT004-0001 × G7NT005-0001 > 04TDNT000007

[0211] G7NT004-0002 × G7NT005-0001 > (无后代)

[0212] G7NT004-0003 × G7NT005-0001 > 04TDNT000012

[0213] G7NT004-0001 × G7NT005-0002 > 04TDNT000008

[0214] G7NT004-0002×G7NT005-0002 > 04TDNT000010

[0215] G7NT004-0003×G7NT005-0002 > 04TDNT000011

[0216] 从每一杂交中,将 200 粒种子播种在 Km(200mg/L) 培养基上、将 200 粒种子播种在 Hyg(50mg/L) 培养基上并且将 200 粒种子播种在 Km(200mg/L)+Hyg(50mg/L) 培养基上以检验转基因的正常传递。对于大多数杂交而言存在相当正常的不同转基因的转移(注意对于某些杂交,遇到污染问题和种子质量问题(见下表)):

[0217] 对相应选择剂抵抗的幼苗数

系	幼苗数/50 粒种子	Km ^R 幼苗数 /200 粒种子	Hyg ^R 幼苗数 /200 粒种子	km ^R +Hyg ^R 幼 苗数/200 粒 种子
G7NT001-0001 × G7NT005-0001	32	47/150	55	28/150
G7NT001-0001 × G7NT005-0002	32	29	51	15
G7NT001-0002 × G7NT005-0001	32	89	64	59
G7NT001-0002 × G7NT005-0002	46	69	94	42
[0218] G7NT001-0003 × G7NT005-0001	47	92	93	53
G7NT001-0003x G7NT005-0002	48	88	85	47
G7NT004-0001 × G7NT005-0001	49	92	65/150	44
G7NT004-0002 × G7NT005-0001	47	73/150	89	34/150
G7NT004-0002 × G7NT005-0002	49	58/150	98	60
G7NT004-0003 × G7NT005-0001	39	63	69	50
G7NT004-0003 × G7NT005-0002	45	60	91	22

[0219] 对于这 12 种杂交中的每一种,将少量 Km^R+Hyg^R 后代植物转移运至温室以作为 WT SR1 植物的授粉者使用。对于这些 12 种杂交,每次使用三株 Km^R+Hyg^R 植物作为根据如下方案的 WT SR1 授粉者:

[0220] SR1×04TDNT000001-001

[0221] -002

[0222] -003

[0223] SR1×04TDNT000002-001

[0224] -002

[0225] -003

[0226] SR1×04TDNT000003-001

[0227] -002

[0228] -003

[0229] SR1×04TDNT000004-001

[0230]	-002
[0231]	-003
[0232]	SR1 × 04TDNT000005-001
[0233]	-002
[0234]	-003
[0235]	SR1 × 04TDNT000006-001
[0236]	-002
[0237]	-003
[0238]	SR1 × 04TDNT000007-001
[0239]	-002
[0240]	-003
[0241]	SR1 × 04TDNT000012-001
[0242]	-002
[0243]	-003
[0244]	SR1 × 04TDNT000008-001
[0245]	-002
[0246]	-003
[0247]	SR1 × 04TDNT000010-001
[0248]	-002
[0249]	-003
[0250]	SR1 × 04TDNT000011-001
[0251]	-002
[0252]	-003

[0253] 从这些杂交的每一后代中（见下表），将 50 粒种子播种在非选择性基质上以测定萌发频率，将 50 粒种子播种在卡那霉素基质上以测定 NTM19-I-SceI 基因的传递率并且将约 4000 粒种子播种在 PPT 基质上以测定从一个世代转移至另一世代期间的 IHR 频率。也呈现 Km^R 的 PPT^R 幼苗数决定着 NTM19-ISceI 内切核酸酶所诱导的 DSB 是否影响从一个世代转至另一世代期间的 IHR 频率。

[0254] 对 22 个后代的后代分析结果汇总于表 A、B 和 C 内。

[0255] NTM19-I-SceI 对从一个世代转移至另一世代期间的 IHR 频率存在极强烈的影响，因为全部 PPT^R 幼苗也呈 Km^R 。

[0256] 应当指出大部分的 PPT^R 和 GFP^F 幼苗没有进一步发育成植物并且因 GFP 的毒性效应而死亡。

[0257]

表 A:

杂交	萌发频率(幼苗数/50 粒种子)	Km ^R 幼苗数/50 粒种子	PPT ^R 和GFP ^F 幼苗数/种子数	Km ^R 幼苗数/针对Km ^R 筛选的PPT ^R 和GFP ^F 幼苗数
SR1×04TDNT000001-001 短重复序列	43	24	77/4348 (1.77%)	5/5
SR1×04TDNT000001-002 短重复序列	49	20	79/4835 (1.63%)	23/23
SR1×04TDNT000001-003 短重复序列	47	22	98/4827 (2.03%)	27/27
SR1×04TDNT000002-001 短重复序列	47	23	33/4762 (0.69%)	4/4
SR1×04TDNT000004-001 短重复序列	49	30	123/4798 (2.6%)	36/36
SR1×04TDNT000004-002 短重复序列	48	23	100/4745 (2.1%)	32/32
SR1×04TDNT000004-003 短重复序列	48	15	118/4665 (2.5%)	6/6
SR1×04TDNT000005-001 短重复序列	49	25	94/4665 (2.01%)	16/16
SR1×04TDNT000005-002 短重复序列	48	20	47/4690 (1%)	7/7
SR1×04TDNT000005-003 短重复序列	48	22	120/4658 (2.6%)	16/18 (2 个为 S 或 R?)
SR1×04TDNT000006-001 短重复序列	47	28	136/4665 (2.9%)	24/24
SR1×04TDNT000006-003 短重复序列	49	20	77/4650 (1.66%)	12/12

[0258]

表 B:

杂交	萌发频率(幼苗数/50 粒种子)*	Km ^R 幼苗数/50 粒种子	Hyg ^R 幼苗数/50 粒种子	Km ^R +Hyg ^{R+} 幼苗数/100 粒种子	PPT ^R 和 GFP ^F 幼苗数/种子数**	Km ^R 幼苗数/针对 Km ^R 筛选的 PPT ^R 和 GFP ^F 幼苗
SR1×04TDNT000003-001 短重复序列	23	14	12	13	44/4973 (0.89%)**	33/33
SR1×04TDNT000003-003 短重复序列	20	16	11	16	46/4857 (0.95%)**	46/46
SR1×04TDNT000007-001 长重复序列	19	7	7	7	16/4915 (0.33%)**	16/16
SR1×04TDNT000008-001 长重复序列	28	17	12	12	33/4890 (0.7%)**	33/33
SR1×04TDNT000008-003 长重复序列	20	7	8	8	33/4840 (0.69%)**	33/33
SR1×04TDNT000012-003 长重复序列	16	10	9	9	14/4312 (0.32%)**	14/14

*同时播种本表内提到的后代。由于漂白剂过度强烈的灭菌作用,存在不良的和不规则的萌发(对于大多数系<50%)。**意味着 PPT^R 和 GFP^F 幼苗数/种子数低估至少因子 2, 因为大多数系的萌发频率小于 50%。

[0259]

表 C:

杂交	萌发频率 (幼苗数/50 粒种子) ^a	Km ^R 幼苗 数/ 50 粒种子	Hyg ^R 幼 苗数 /50 粒种子	Km ^R 幼 +Hyg ^R 幼 苗数 /100 粒种子	PPT ^R 和GFP ^F 幼苗种子数**	Km ^R 幼苗数/ 针对Km ^R 筛选 的 PPT ^R 和 GFP ^F 幼苗数 ^a
SR1×04TDNT000002-002 短重复序列	50	20	26	9	7/1330 (0.5%)	NT ^a
SR1×04TDNT000002-003 短重复序列	50	30	18	25	9/1355 (0.66%)	NT ^a
SR1×04TDNT000003-002 短重复序列	50	20	21	25	24/1389 (1.7%)	NT ^a
SR1×04TDNT000007-003 长重复序列	50	25	25	17	3/1346 (0.2%)	NT ^a

^aNT: 未测试

[0260] 此外,全部 PPT^R 和 GFP^F 幼苗确实是潮霉素敏感的,这表明 hyg 基因的确已经通过在 IHR 基因座内的染色体间重组被去除。

[0261]

杂交	Hyg ^R 幼苗数/对 Hyg ^R 筛选的 PPT ^R 和 GFP ^F 幼苗数
SR1×04TDNT000012-003	0/11
SR1×04TDNT000008-001	0/12
SR1×04TDNT000008-003	0/11
SR1×04TDNT000001-002	0/8
SR1×04TDNT000005-003	0/7
SR1×04TDNT000006-003	0/7

[0262] 从 18 个后代群体的分离分析中,可以得出结论,NTM19-I-SceI 对从一个世代转移至另一世代期间的 IHR 频率存在极强烈的影响,因为全部 PPT^R 幼苗也呈 Km^R。

[0263] SR1 (雌性) 与 04TDNT00000X-00Y 间的杂交的后代通常将分离成:

[0264] 25% 仅具有 NTM19-I-SceI 内切核酸酶

[0265] 25% 仅具有 IHR 构建体

[0266] 25% 同时具有 NTM19-I-SceI 内切核酸酶 + IHR 构建体

[0267] 25% 既没有 NTM19-I-SceI 内切核酸酶,也没有 IHR 构建体

[0268] 全部 PPT^R 幼苗也呈 Km^R 的事实表明,全部 IHR 重组仅发生在含有处于 NTM19 小孢子特异性启动子控制下的 I-SceI 内切核酸酶和 IHR 构建体的部分内。我们的结果表明,在最佳情况下多达 11% 的含有 NTM19-I-SceI 内切核酸酶 + IHR 构建体的小孢子已经发生了导致缺陷性 egfp-bar 融合基因恢复的染色体同源重组 (SR1×04TDNT000006-001)。由于在仅含有 IHR 构建体的部分中没有得到产生功能性 egfp-bar 基因的 IHR 重组,我们可以得出结论,即自发性 IHR (在小孢子内缺乏定向 DSB 诱导情况下) 没有发生,或如果自发性 IHR 的确发生,则它没有导致缺陷性 egfp-bar 融合基因恢复。相反,小孢子内 DSB 诱导的 IHR 允许导致缺陷性 egfp-bar 融合基因恢复的更精确的染色体同源重组。

[0269] 序列分析显示在通过小孢子内 DSB 诱导的 IHR 介导去除选择标记后,没有留下痕迹。

[0001]

序列表

<110> 拜尔生物科学公司(BAYER BIOSCIENCE N.V.)

<120> 用于去除所选择 DNA 序列的方法和手段

<130> BCS 05-2007

<150> EP05075781.4

<151> 2005-04-04

<150> US60/669,243

<151> 2005-04-07

<160> 6

<170> PatentIn 3.3 版

<210> 1

<211> 732

<212> DNA

<213> 人工的

<220>

<223> 合成的 I-SceI 编码区(UIPAC)

<220>

<221> 变异

<222> (25)..(27)

<223> AGA

<220>

<221> 变异

<222> (73)..(75)

<223> AGC

<220>

<221> 变异

<222> (97)..(99)

<223> AGC

<220>

[0002]

<221> 变异
<222> (169).. (171)
<223> AGA

<220>
<221> 变异
<222> (172).. (174)
<223> AGC

<220>
<221> 变异
<222> (175).. (177)
<223> AGA

<220>
<221> 变异
<222> (268).. (270)
<223> AGC

<220>
<221> 变异
<222> (289).. (291)
<223> AGA

<220>
<221> 变异
<222> (436).. (438)
<223> AGC

<220>
<221> 变异
<222> (490).. (492)
<223> AGC

<220>
<221> 变异
<222> (502).. (504)
<223> AGC

<220>
<221> 变异
<222> (523).. (525)
<223> AGC

[0003]

<220>

<221> 变异

<222> (565).. (567)

<223> AGA

<220>

<221> 变异

<222> (631).. (633)

<223> AGC

<220>

<221> 变异

<222> (637).. (639)

<223> AGC

<220>

<221> 变异

<222> (712).. (714)

<223> AGC

<220>

<221> 变异

<222> (715).. (717)

<223> AGC

<400> 1

atggcyaarc chcchaaraa raarcgsaaa gtsaacatya araaraacca ggtsatgaac 60

ctsggmcccha actcmaarct sctsaargag tacaartcmc arctsatyga rctsaacaty 120

garcarttcg argcyggmat cggmctsaty ctsggmgayg cytacatygc stcmcgsgay 180

garggmaara cytactgyat gcagttcgar tggaaraaca argcytacat ggaycaygts 240

tgyctstctst acgaycartg ggtstctstcm cchcchcaya araargarcg sgtsaaccay 300

ctsggmaacc tsqtsatyac ytggggmgy caracyttca arcaycargc yttcaacaar 360

ctsgcsaacc tttcatyct saacaacaar aaracyatyc chaacaacct sgtsgaraac 420

tacctsacyc cyatgtcmct sgcytactgg ttcatggayg ayggmgmaa rtgggaytac 480

aacaaraact cmacyaaca rctmatygtts ctsaacacyc artcmittcac yttcgargar 540

[0004]

gtsgartacc tsgtsaargg mctscgsaac aarttccarc tsaactgyta cgtsaagaty	600
aacaaraaca arccyatyat ctacatygay tcmatgtcmt acctsatytt ctacaaccts	660
atyaarccht acctsatycc hcaratgatg tacaarctsc chaacacyat ytcmtcmgar	720
acyttcctsa ar	732
<210> 2	
<211> 732	
<212> DNA	
<213> 人工的	
<220>	
<223> 合成的 I-SceI 编码区	
<400> 2	
atggccaagc ctccaagaa gaagcgcaaa gtgaacatca agaagaacca ggtgatgaac	60
ctgggaccta acagcaagct cctgaaggag tacaagagcc agctgatcga actgaacatc	120
gagcagttcg aagctggcat cggcctgatc ctgggcgatg cctacatcag atcccgggac	180
gaaggcaaga cctactgcat gcagttcgag tggaagaaca aggcctacat ggaccacgtg	240
tgtctgctgt acgaccagtg ggtcctgagc cctcctcaca agaaggagcg cgtgaaccat	300
ctgggcaacc tcgtgatcac ctggggagcc cagaccttca agcaccaggc cttcaacaag	360
ctggccaacc tgttcacgt gaacaacaag aagaccatcc ccaacaacct cgtggagaac	420
tacctcactc ccatgagcct ggcctactgg ttcattggacg acggaggcaa gtgggactac	480
aacaagaaca gcaccaacaa gtcaattgtg ctgaacaccc aaagcttcac cttcgaagaa	540
gtggagtacc tcgtcaaggc cctgcgcaac aagttccagc tgaactgcta cgtgaagatc	600
aacaagaaca agcctatcat ctacatcgac agcatgagct acctgatctt ctacaacctg	660
atcaagccat acctgatccc tcagatgatg tacaagctgc ccaacacat cagcagcgag	720
accttcctga ag	732

[0005]

<210> 3
<211> 1618
<212> DNA
<213> 烟草(Nicotiana tabacum)

<220>
<221> TATA_信号
<222> (919)..(922)

<220>
<221> mRNA
<222> (954)..(1573)

<220>
<221> 基因
<222> (993)..(1271)
<223> NTM19

<220>
<221> misc_feature
<222> (993)..(1271)
<223> 编码区

<400> 3
gatccagatt tatagggtcc taatgcgggt actgaacacc aggtgggaaa caaaaaatat 60
acagaacaac tccttttagaa tttacaattt ttgagcgtgt tggcttggtta cgattctact 120
tttcatatct ctcgtcatct cctaactcct atggttcacc agccaccgat taattatgac 180
accgctaaca aaaatcttgc gacgacattg agagaaattt cttttcataa attggttaatt 240
cgtacatcat ttataggcgt tagctataac ctttttagtta gtgaatacaa tactttttgc 300
tattattatg taacttttag atatgaattt actttcaaaa aaaaaaaaaag gatcgaattt 360
ggttatcaac taaggaccaa ccactttgga cgtctcacca ctaagttaa taaatcactt 420
tgttctcgaa aaaaacccca aaagtgttaa aatgcttttc atatcataat caaacaacgt 480
gattaataaa atctattaag ttaatagaag tagggaataa atcgggcaaa agaatttgat 540

[0006]

acaaacaaa cgggtcaaaa aagctagtat tcatataaat ggactataca agttaatacc	600
agctagcaga aattaaatag tttattaagt tgattacaaa acaattcctc atttaaaaaa	660
agttaatgta atcaagagat cttttgcttc taattgatca gacgaggacc cctcttattt	720
atcttctttt tcatataaga ttttgaatag atataggga atcttgttca ctctttatct	780
acttcaaatt gcatgcattt taagaattct ctttgtatgc aaacttcagt atttatgatt	840
gacataaatc aatattcata tcttcgataa agttaataac tctcctaata cttatgaata	900
tctcttcctt tacaacccta taaaaccccc cactatagct accctcataa ttcattcttag	960
agtaccaacc ctaaatttct tagtgattaa ccatggctaa gaaaagtctc acttttctca	1020
tttgcatttt cctaattctc aatttatgtt ttgcaattga gaacgtagaa actatgcaaa	1080
aatcggattc atcgtcacia gataaagaat tagattgggt tcatccgtgg ttccatccac	1140
atccatgggt gctacatcca catccatggc cattcggttca tccgccaatg ccagctggcg	1200
gttttcatca tgcattggcca ttcccccatc caccgatgcc tgctgggtgt ttttaagttc	1260
ctcatcaata atttcatcgt catccatggc cattcatgca tccaccagtt ccattctccac	1320
ctaaaggaga caagaattaa ttgaaaatat gaagagaagt gttggatcaa catcttattg	1380
atcacatatt tttctttagg ttaatatctt taggatttat gtcttaggtt atttttgata	1440
aaaattaaaa taaatgatcg ttctagggtg gttattataa tttcttagat ttttccaagt	1500
agctttcgat ggtagaatg ttattaattt gattcggctt atcatgaaat aaaatccgta	1560
gtattattgc tttagctttt atgatttgta gttattttat gttgattgtt ctccattt	1618

<210> 4

<211> 4683

<212> DNA

<213> 人工的

<220>

<223> pTCV63 的 T-DNA

[0007]

<220>
<221> misc_feature
<222> (191)..(222)
<223> T-DNA 右边界

<220>
<221> misc_feature
<222> (270)..(818)
<223> CaMV35S 启动子区域

<220>
<221> misc_feature
<222> (819)..(875)
<223> Cab22 前导序列

<220>
<221> misc_feature
<222> (885)..(1391)
<223> eGFP 编码区的 5' 部分

<220>
<221> misc_feature
<222> (1392)..(1409)
<223> I-SceI 识别位点

<220>
<221> misc_feature
<222> (1411)..(1419)
<223> Zif268 识别位点

<220>
<221> misc_feature
<222> (1433)..(1671)
<223> 3' 35S 聚腺苷酸化区域(互补序列)

<220>
<221> misc_feature
<222> (1683)..(2708)
<223> 潮霉素抗性编码区(互补序列)

<220>
<221> misc_feature

[0008]

<222> (2715).. (2787)
 <223> CsVMV 前导区(互补序列)

<220>
 <221> misc_feature
 <222> (2788).. (3227)
 <223> CsVMV 启动子片段(互补序列)

<220>
 <221> misc_feature
 <222> (3269).. (3787)
 <223> eGFP 的 3' 部分

<220>
 <221> misc_feature
 <222> (3788).. (4339)
 <223> 磷丝菌素乙酰转移酶编码区

<220>
 <221> misc_feature
 <222> (4341).. (4549)
 <223> 3' nos:转录终止和聚腺苷酸化区域

<220>
 <221> misc_feature
 <222> (4659).. (4683)
 <223> T-DNA 右边界

<400> 4
 agattcgaag ctcggtcccg tgggtgttct gtcgtctcgt tgtacaacga aatccattcc 60
 cattccgcgc tcaagatggc ttcccctcgg cagttcatca gggctaaatc aatctagccg 120
 acttgtccgg tgaaatgggc tgcactccaa cagaaacaat caaacaacaa tacacagcga 180
 cttattcaca cgcgacaaat tacaacggta tatatcctgc cagtactcgg ccgtcgacct 240
 gcaggcaatt ggtacctacg tatgcatggc gcgccatatg caccatacat ggagtcaaaa 300
 attcagatcg aggatctaac agaactcgcc gtgaagactg gcgaacagtt catacagagt 360
 cttttacgac tcaatgacaa gaagaaaatc ttcgtcaaca tgggtggagca cgacactctc 420
 gtctactcca agaatatcaa agatacagtc tcagaagacc aaagggttat tgagactttt 480

[0009]

caacaaaggg taatatcggg aaacctctc ggattccatt gccagctat ctgtcacttc	540
atcaaaagga cagtagaaaa ggaagggtgc acctacaaat gccatcattg cgataaagga	600
aaggctatcg ttcaagatgc ctctgccgac agtgggtccca aagatggacc cccaccacg	660
aggagcatcg tggaaaaaga agacgttcca accacgtctt caaagcaagt ggattgatgt	720
gatatctcca ctgacgtaag ggatgacgca caatcccact atccttcgca agacccttcc	780
tctatataag gaagttcatt tcatttggag aggactcgag ctcatctctc tattacttca	840
gccataacaa aagaactctt ttctcttctt attaaaccaa aaccatggtg agcaagggcg	900
aggagctgtt caccgggggtg gtgcccatcc tggtcgagct ggacggcgac gtaaacggcc	960
acaagttcag cgtgtccggc gagggcgagg gcgatgccac ctacggcaag ctgaccctga	1020
agttcatctg caccaccggc aagctgcccg tgccctggcc caccctcgtg accaccctga	1080
cctacggcgt gcagtgttc agccgtacc cgcaccacat gaagcagcac gacttcttca	1140
agtccgccat gccgaaggc tacgtccagg agcgcacat cttcttcaag gacgacggca	1200
actacaagac ccgcgccgag gtgaagttcg agggcgacac cctggtgaac cgcacgagc	1260
tgaagggcat cgacttcaag gaggacggca acatcctggg gcacaagctg gagtacaact	1320
acaacagcca caacgtctat atcatggccg acaagcagaa gaacggcatc aaggtgaact	1380
tcaagatccg ctagggataa cagggtaatg gcgtgggcga ccggtcccgg ccgctgtacc	1440
atgcatgac tggattttag tactggattt tggttttagg aattagaaat tttattgata	1500
gaagtatittt acaatacaa atacatacta agggtttctt atatgctcaa cacatgagcg	1560
aaaccctata ggaaccctaa ttcccttacc tgggaactac tcacacatta ttatggagaa	1620
aatagagaga gatagatttg tagagagaga ctggtgattt cagcgtgtcc aagcttgata	1680
tcctattcct ttgccctcgg acgagtgtg gggcgtcggg ttccactatc ggcgagtact	1740
tctacacagc catcgggtcca gacggccgcg cttctgcggg cgatttgtgt acgcccgaca	1800

[0010]

gtccccgctc cggatcggac gattgcgtcg catcgaccct gcgccaagc tgcacatcg	1860
aaattgccgt caaccaagct ctgatatagat tggtaagac caatgcggag catatacgcc	1920
cggagccgcg gcgacatctgc aagctccgga tgcctccgct cgaagtagcg cgtctgctgc	1980
tccatacaag ccaaccacgg cctccagaag aagatgttgg cgacctcgta ttgggaatcc	2040
ccgaacatcg cctcgtccca gtcaatgacc gctgttatgc ggccattgtc cgtcaggaca	2100
ttgttggagc cgaaatccgc gtgcacgagg tgccggactt cggggcagtc ctcggcccaa	2160
agcatcagct catcgagagc ctgcgcgacg gacgcactga cgggtgtcgtc catcacagtt	2220
tgccagtgat acacatgggg atcagcaatc gcgcataatga aatcacgcca tgtagtgtat	2280
tgaccgattc cttgcgggtcc gaatgggccc aaccgcctcg tctggctaag atcgcccgca	2340
gcgacgcat ccatggcctc cgcgaccggc tgcagaacag cgggcagttc ggtttcaggc	2400
aggtcttgca acgtgacacc ctgtgcacgg cgggagatgc aataggtcag gctctcgctg	2460
aattccccaa tgtcaagcac ttccggaatc gggagcgcgg ccgatgcaa gtgccgataa	2520
acataacgat cttttagtaa accatcggcg cagctattta cccgcaggac atatccacgc	2580
cctcctacat cgaagctgaa agcacgagat tcttcgcctt ccgagagctg catcaggtcg	2640
gagacgtgt cgaacttttc gatcagaaac ttctcgacag acgtcgcggt gagttcaggc	2700
tttttcatct cgagacaaac ttacaaattt ctctgaagtt gtatcctcag tacttcaaag	2760
aaaatagctt acaccaaatt ttttcttgtt ttacaaaatg ccgaacttgg ttccttatat	2820
aggaaaactc aagggcaaaa atgacacgga aaaatataaa aggataagta gtgggggata	2880
agattccttt gtgataaggt tactttccgc ccttacattt tccaccttac atgtgtcctc	2940
tatgtctctt tcacaatcac cgaccttate ttcttctttt cattgttgte gtcagtgtt	3000
acgtcttcaa gattcttttc ttgccttgtt tcttctttt caatttctac gtattcttct	3060
tcgtattctg gcagtatagg atcttgatc tgtacattct tcatttttga acataggttg	3120

[0011]

catatgtgcc gcatattgat ctgcttcttg ctgagctcac ataatacttc catagttttt	3180
cccgtaaaca ttggattctt gatgctacat cttggataat taccttcggc gcgccaatgca	3240
tacgtaggta ccaattgccg ggaccggtta cggcgtgcag tgcttcagcc gctaccccga	3300
ccacatgaag cagcacgact tcttcaagtc cgccatgccc gaaggctacg tccaggagcg	3360
caccatcttc ttcaaggacg acggcaacta caagaccgc gccgaggtga agttcgaggg	3420
cgacaccctg gtgaaccgca tcgagctgaa gggcatcgac ttcaaggagg acggcaacat	3480
cctggggcac aagctggagt acaactacaa cagccacaac gtctatatca tggccgacaa	3540
gcagaagaac ggcatcaagg tgaacttcaa gatccgccac aacatcgagg acggcagcgt	3600
gcagctcgcc gaccactacc agcagaacac ccccatcggc gacggccccg tgctgctgcc	3660
cgacaaccac tacctgagca ccagtcgc cctgagcaaa gacccaacg agaagcgca	3720
tcacatggtc ctgctggagt tcgtgaccgc cgccgggatac actctcggca tggacgagct	3780
gtacaagatg gaccagaac gacggccgc cgacatccgc cgtgccaccg agggcgacat	3840
gccggcggtc tgcacatcg tcaaccacta catcgagaca agcacggtca acttcgtac	3900
cgagccgcag gaaccgcagg agtggacgga cgacctcgtc cgtctgcggg agcgctatcc	3960
ctggctcgtc gccgaggtgg acggcgaggt cgccggcatac gcctacgcgg gccctggaa	4020
ggcacgcaac gcctacgact ggacggccga gtcgaccgtg tacgtctccc cccgccacca	4080
gcggacggga ctgggctcca cgctctacac ccacctgctg aagtccttg aggcacagg	4140
cttcaagagc gtggtcgtg tcatcgggct gccaacgac ccgagcgtgc gcatgcacga	4200
ggcgctcgga tatgcccc gcgcatgct gcggcgggc ggcttcaagc acgggaactg	4260
gcatgacgtg ggtttctggc agctggactt cagcctgccg gtaccgccc gtccggtcct	4320
gcccgtcacc gagatctgag ctagcacgcg tctaggatcc gaagcagatc gttcaaact	4380
ttggcaataa agtttcttaa gattgaatcc tgttccggc cttgcgatga ttatcatata	4440

[0012]

atttctgttg aattacgtta agcatgtaat aattaacatg taatgcatga cgttatttat 4500
gagatgggtt tttatgatta gagtcccgca attatacatt taatacgca tagaaaacaa 4560
aatatagcgc gcaaaactagg ataaattatc gcgcgcggtg tcatttatgt tactagatcg 4620
ggaagatcct ctagatacgt agcgatcgcc atggagccat ttacaattga atatatcctg 4680
ccg 4683

<210> 5

<211> 4992

<212> DNA

<213> 人工的

<220>

<223> pTCV64 的 T-DNA

<220>

<221> misc_feature

<222> (191)..(222)

<223> T-DNA 右边界

<220>

<221> misc_feature

<222> (270)..(818)

<223> CaMV35S 启动子片段

<220>

<221> misc_feature

<222> (819)..(875)

<223> cab22 前导片段

<220>

<221> misc_feature

<222> (885)..(1553)

<223> eGFP 编码区的 5' 部分

<220>

<221> misc_feature

<222> (1554)..(1571)

[0013]

- <223> I-SceI 识别位点
- <220>
- <221> misc_feature
- <222> (1573).. (1581)
- <223> Zif268 识别位点
- <220>
- <221> misc_feature
- <222> (1595).. (1833)
- <223> 3' 35S 转录终止和聚腺苷酸化区域(互补序列)
- <220>
- <221> misc_feature
- <222> (1845).. (2870)
- <223> 潮霉素磷酸转移酶编码区
- <220>
- <221> misc_feature
- <222> (2877).. (2949)
- <223> CsVMV 的前导区(互补序列)
- <220>
- <221> misc_feature
- <222> (2950).. (3389)
- <223> CsVMV 启动子片段(互补序列)
- <220>
- <221> misc_feature
- <222> (3431).. (4096)
- <223> eGFP 编码区的 3' 部分
- <220>
- <221> misc_feature
- <222> (4097).. (4648)
- <223> 磷丝菌素乙酰转移酶编码区
- <220>
- <221> misc_feature
- <222> (4650).. (4858)
- <223> 3' nos: 胭脂氨酸合酶基因的转录终止和聚腺苷酸化区域
- <220>
- <221> misc_feature

[0014]

<222> (4968).. (4992)

<223> T-DNA 左边界序列

<400> 5

```

agattcgaag ctccgtcccg tgggtgttct gtcgtctcgt tgtacaacga aatccattcc      60

cattccgcgc tcaagatggc ttcccctcgg cagttcatca gggctaaatc aatctagccg      120

acttgtccgg tgaaatgggc tgcactccaa cagaaacaat caaacaacaa tacacagcga      180

cttattcaca cgcgacaaat tacaacggta tatatcctgc cagtactcgg ccgtcgacct      240

gcaggcaatt ggtacctacg tatgcatggc gcgccatatg caccatacat ggagtcaaaa      300

attcagatcg aggatctaac agaactcgcc gtgaagactg gcgaacagtt catacagagt      360

cttttacgac tcaatgacaa gaagaaaatc ttcgtcaaca tgggtggagca cgacactctc      420

gtctactcca agaatatcaa agatacagtc tcagaagacc aaagggtat tgagactttt      480

caacaaaggg taatatcggg aaacctcctc ggattccatt gccagctat ctgtcacttc      540

atcaaaagga cagtagaaaa ggaaggtggc acctacaaat gccatcattg cgataaagga      600

aaggctatcg ttcaagatgc ctctgccgac agtgggtcca aagatggacc cccaccacg      660

aggagcatcg tggaaaaaga agacgttcca accacgtctt caaagcaagt ggattgatgt      720

gatatctcca ctgacgtaag ggatgacgca caatcccact atccttcgca agacccttcc      780

tctatataag gaagttcatt tcatttggag aggactcgag ctcatctctc tattacttca      840

gccataacaa aagaactctt ttctcttctt attaaacaa aaccatgggtg agcaagggcg      900

aggagctgtt caccgggggtg gtgcccaccc tggctcgagct ggacggcgac gtaaacggcc      960

acaagttcag cgtgtccggc gagggcgagg gcgatgccac ctacggcaag ctgaccctga     1020

agttcatctg caccaccggc aagctgcccc tgccctggcc caccctcgtg accaccctga     1080

cctacggcgt gcagtgttc agccgctacc ccgaccacat gaagcagcac gacttcttca     1140

agtcgcccat gcccgaaggc tacgtccagg agcgcaccat cttcttcaag gacgacggca     1200

```

[0015]

actacaagac cgcgcgcgag gtgaagttcg agggcgacac cctggtgaac cgcacgcgagc	1260
tgaagggcat cgacttcaag gaggacggca acatcctggg gcacaagctg gagtacaact	1320
acaacagcca caacgtctat atcatggccg acaagcagaa gaacggcatc aaggtgaact	1380
tcaagatccg ccacaacatc gaggacggca gcgtgcagct cgccgaccac taccagcaga	1440
acacccccat cggcgacggc cccgtgtctg tgcccagaaa ccactacctg agcaccacgt	1500
ccgccctgag caaagacccc aacgagaagc gcgacacat ggtcctgtcg gtagtaggat	1560
aacagggtaa tggcgtgggc gaccgggtccc ggccgctgta ccatgcatga tctggatttt	1620
agtactggat tttggtttta ggaattagaa attttattga tagaagtatt ttacaaatac	1680
aaatacatac taagggtttc ttatatgtc aacacatgag cgaaacccta taggaaccct	1740
aattccctta tctgggaact actcacacat tattatggag aaaatagaga gagatagatt	1800
tgtagagaga gactggtgat ttcagcgtgt ccaagcttga taccctattc ctttgcctc	1860
ggacgagtgc tggggcgctg gtttccacta tcggcgagta cttctacaca gccatcggtc	1920
cagacggccg cgcttctgcg ggcgatttgt gtacggccga cagtcgccgc tccggatcgg	1980
acgattgcgt cgcacgcacc ctgcgcccga gctgcatcat cgaaattgcc gtcaaccaag	2040
ctctgataga gttggtcaag accaatgcgg agcatatacg cccggagccg cggcgatcct	2100
gcaagctccg gatgcctccg ctggaagtag cgcgtctgct gctccataca agccaaccac	2160
ggcctccaga agaagatgtt ggcgacctcg tattgggaat ccccgaacat cgcctcgctc	2220
cagtcaatga ccgctgttat gcggccattg tccgtcagga cattgttgga gccgaaatcc	2280
gcgtgcacga ggtgccggac ttcggggcag tcctcgggcc aaagcatcag ctcatcgaga	2340
gcctgcgcga cggacgcact gacggtgtcg tccatcacag ttgcccagtg atacacatgg	2400
ggatcagcaa tcgcgcatat gaaatcacgc catgtagtgt attgaccgat tccttgcggt	2460
ccgaatgggc cgaacccgct cgtctggcta agatcgggcc cagcgcgcgc atccatggcc	2520

[0016]

tccgcgaccg gctgcagaac agcgggcagt tcggtttcag gcaggtcttg caacgtgaca	2580
ccctgtgcac ggcgggagat gcaataggtc aggcctctgc tgaattcccc aatgtcaagc	2640
acttccggaa tcgggagcgc ggccgatgca aagtgccgat aaacataacg atctttgtag	2700
aaaccatcgg cgcagctatt taccgcagg acatatccac gccctcctac atcgaagctg	2760
aaagcacgag attcttcgcc ctccgagagc tgcacaggt cggagacgct gtcgaacttt	2820
tcgatcagaa acttctcgac agacgtcgcg gtgagttcag gctttttcat ctcgagacaa	2880
acttacaaat ttctctgaag ttgtatctc agtacttcaa agaaaatagc ttacacaaaa	2940
ttttttcttg ttttcacaaa tgccgaactt ggttccttat ataggaaaac tcaagggcaa	3000
aatgacacg gaaaaatata aaaggataag tagtggggga taagattcct ttgtgataag	3060
gttactttcc gcccttacat ttccacctt acatgtgtcc tctatgtctc tttcacaatc	3120
accgacctta tcttcttctt ttcatgttg tcgtcagtc ttacgtcttc aagattcttt	3180
tcttcgctg gttcttcttt ttcaatttct acgtattctt cttcgtattc tggcagtata	3240
ggatcttgta tctgtacatt cttcatTTTT gaacataggt tgcataatgt ccgcataattg	3300
atctgcttct tgetgagctc acataatact tccatagttt ttcccgtaaa cattggattc	3360
ttgatgctac atcttgata attaccttcg gcgcgccatg catacgtagg taccaattgc	3420
cgggaccggt gagctggacg gcgacgtaaa cggccacaag ttcagcgtgt ccggcgaggg	3480
cgagggcgat gccacctacg gcaagctgac cctgaagttc atctgcacca ccggcaagct	3540
gcccgtgccc tggcccaccc tcgtgaccac cctgacctac ggctgcagt gcttcagccg	3600
ctaccccgac cacatgaagc agcacgactt cttcaagtcc gccatgcccg aaggctacgt	3660
ccaggagcgc accatcttct tcaaggacga cggcaactac aagaccgcg ccgaggtgaa	3720
gttcgagggc gacaccctgg tgaaccgcat cgagctgaag ggcatcgact tcaaggagga	3780
cggcaacatc ctggggcaca agctggagta caactacaac agccacaacg tctatatcat	3840

[0017]

ggccgacaag cagaagaacg gcatcaaggt gaacttcaag atccgccaca acatcgagga	3900
cggcagcgtg cagctcgccg accactacca gcagaacacc cccatcggcg acggccccgt	3960
gctgctgccc gacaaccact acctgagcac ccagtcgcc ctgagcaaag accccaacga	4020
gaagcgcgat cacatggtcc tgctggagtt cgtgaccgcc gccgggatca ctctcgcat	4080
ggacgagctg tacaagatgg acccagaacg acgcccggcc gacatccgcc gtgccaccga	4140
ggcggacatg ccggcggtct gcaccatcgt caaccactac atcgagacaa gcacggtcaa	4200
cttccgtacc gagccgcagg aaccgcagga gtggacggac gacctcgtcc gtctcgggga	4260
gcgctatccc tggctcgtcg ccgaggtgga cggcgaggtc gccggcatcg cctacgcggg	4320
cccctggaag gcacgcaacg cctacgactg gacggccgag tcgaccgtgt acgtctcccc	4380
ccgccaccag cggacgggac tgggtccac gctctacacc cacctgctga agtcacctga	4440
ggcacagggc ttcaagagcg tggtcgctgt catcgggctg cccaacgacc cgagcgtgcg	4500
catgcacgag gcgctcggat atgcccccg cggcatgctg cgggcggccg gtttcaagca	4560
cgggaactgg catgacgtgg gtttctggca gctggacttc agcctgccgg taccgccccg	4620
tccggtcctg ccggtcaccg agatctgagc tagcacgcgt ctaggatccg aagcagatcg	4680
ttcaaacatt tggcaataaa gtttcttaag attgaatcct gttgccggtc ttgcgatgat	4740
tatcatataa tttctgttga attacgttaa gcatgtaata attaacatgt aatgcatgac	4800
gttatattatg agatgggttt ttatgattag agtcccgcaa ttatacattt aatacgcat	4860
agaaaacaaa atatagcgcg caaactagga taaattatcg cgcgcggtgt catctatgtt	4920
actagatcgg gaagatcctc tagatacgta gcgatgccca tggagccatt tacaattgaa	4980
tatatcctgc cg	4992

<210> 6

<211> 3987

<212> DNA

[0018]

<213> 人工的

<220>

<223> pTCV72 的 T-DNA

<220>

<221> misc_feature

<222> (191).. (222)

<223> T-DNA 右边界

<220>

<221> misc_feature

<222> (251).. (539)

<223> 3' ocs: 章鱼碱合酶基因的聚腺苷酸化区域和聚腺苷酸化区域 (互补序列)

<220>

<221> misc_feature

<222> (713).. (1510)

<223> 新霉素磷酸转移酶 II 编码区 (互补序列)

<220>

<221> misc_feature

<222> (1511).. (1529)

<223> 接头序列

<220>

<221> misc_feature

<222> (1530).. (1816)

<223> 胭脂氨酸合酶启动子片段 (互补序列)

<220>

<221> misc_feature

<222> (1817).. (1929)

<223> 接头序列

<220>

<221> misc_feature

<222> (1941).. (2926)

<223> 来自烟草的 NTM19 基因的启动子片段

<220>

<221> misc_feature

<222> (2928).. (2961)

[0019]

<223> 核定位信号

<220>

<221> misc_feature

<222> (2962)..(3661)

<223> I-SceI 编码区

<220>

<221> misc_feature

<222> (3665)..(3913)

<223> 3' 35S 转录终止和聚腺苷酸化片段

<220>

<221> misc_feature

<222> (3963)..(3987)

<223> T-DNA 左边界区

<400> 6

```

agattcgaag ctcggtcccg tgggtgttct gtcgtctcgt tgtacaacga aatccattcc      60

cattccgcgc tcaagatggc ttcccctcgg cagttcatca gggctaaatc aatctagccg      120

acttgtccgg tgaatgggc tgcactcaa cagaacaat caaacaaca tacacagcga      180

cttattcaca cgcgacaaat tacaacggtat tatatcctgc cagtactcgg ccgtcgacct      240

gcaggcaatt gtttgttatt gtggcgctct atcatagatg tcgtataaa cctattcagc      300

acaatatatt gttttcattt taatatigta catataagta gtaggttaca atcagtaaatt      360

tgaacggaga atattattca taaaataacg atagtaacgg gtgatataatt cattagaatg      420

aaccgaaacc ggcggtaagg atctgagcta cacatgctca gggttttttac aacgtgcaca      480

acagaattga aagcaaatat catgcgatca taggcgtctc gcatatctca ttaaagcagg      540

gggtgggcga agaactccag catgagatcc ccgcgctgga ggatcatcca gccggcgctcc      600

cggaaaacga ttccgaagcc caacctttca tagaaggcgg cgggtggaatc gaaatctcgt      660

gatggcaggt tgggcgtcgc ttggtcggtc atttcgaacc ccagagtccc gtcagaaga      720

actcgtcaag aaggcgatag aaggcgatgc gctgcgaatc gggagcggcg ataccgtaaa      780

```

[0020]

gcacgaggaa gcggtcagcc cattcgccgc caagctcttc agcaatatca cgggtagcca	840
acgctatgtc ctgatagcgg tccgccacac ccagccggcc acagtcgatg aatccagaaa	900
agcggccatt ttccaccatg atattcggca agcaggcatc gccatgggtc acgacgagat	960
cctcgccgtc gggcatgcgc gccttgagcc tggcgaacag ttcggctggc gcgagcccct	1020
gatgtctctc gtccagatca tcctgatcga caagaccggc ttccatccga gtacgtgtc	1080
gctcgatgcg atgtttcgt tggtagtcga atgggcaggt agccggatca agcgtatgca	1140
gccgccgat tgcacagcc atgatggata ctttctcggc aggagcaagg tgagatgaca	1200
ggagatcctg ccccgccact tcgcccaata gcagccagtc ctttccgct tcagtgacaa	1260
cgtcgagcac agctcgcaa ggaacgccc tcgtggccag ccacgatagc cgcgtgcct	1320
cgtcctgcag ttcatcagg gcaccggaca ggtcggctctt gacaaaaaga accgggcgcc	1380
cctgcgtga cagccggaac acggcggcat cagagcagcc gattgtctgt tgtgccagt	1440
catagccgaa tagcctctcc acccaagcgg ccggagaacc tgcgtgcaat ccatcttggt	1500
caatccacat gatcatgggc cggatctttg attgagagt aatatgagac tctaattgga	1560
taccgagggg aatttatgga acgtcagtg agcatttttg acaagaaata tttgctagct	1620
gatagtgacc ttaggcgact tttgaacgc caataatggt ttctgacgta tgtgcttagc	1680
tcattaaact ccagaaacct gcggtgagt ggctccttca atcgttggg ttctgtcagt	1740
tccaaacgta aaacggcttg tcccgcgtca tcggcggggg tcataacgtg actcccttaa	1800
ttctccgctc atgatcctgt ttctgtgtg aaattgttat ccgctcaca ttccacacat	1860
tatacgagcc ggaagcataa agtgtaaagc ctggggtgcc taatgagtga gaattgacgg	1920
gatctatggc gcgcatatg agatttatag ggtcctaag cgggtactga acaccaggtg	1980
ggaaacaaaa aatatacaga acaactcctt tagaatttac aatttttgag cgtgttggt	2040
tggtacgatt ctacttttca tatctctcgt catctcctaa ctctatggt tcaccagcca	2100

[0021]

ccgattaatt atgacaccgc taacaaaaat cttgcgacga cattgagaga aatttctttt	2160
cataaatggg taattcgtac atcattttata ggcgttagct ataacctttt agttagtgaa	2220
tacaatactt tttgctatta ttatgtaact tttagatatg aatttacttt caaaaaaaaa	2280
aaaaggatcg atgttggtta tcaactaagg accaaccact ttggacgtct caccactaag	2340
ttaaataaat cactttgttc tcgaaaaaaaa ccccaaaagt gttaaaatgc ttttcatatc	2400
ataatcaaac aacgtgatta ataaaatcta ttaagttaat agaagtaggg aataaatcgg	2460
gcaaaagaat ttgatacaaa ccaaaccggt caaaaaagct agtattcata taaatggact	2520
atacaagtta ataccagcta gcagaaatta aatagtttat taagttgatt acaaaacaat	2580
tcctcattta aaaaaagtta atgtaatcaa gagatctttt gcttctaatt gatcagacga	2640
ggacccctct tatttatfff ctttttcata taagattttg aatagatata gggaaatctt	2700
gttcaactct tctctacttc aaattgcatg cattttaaga attctctttg tatgcaaact	2760
tcagtattta tgattgacat aaatcaatat tcatatcttc gataaagtta ataactctcc	2820
taatacttat gaatatctct tcctttacaa ccctataaaa cccccacta tagctacctt	2880
cataattcat cttagagtac caaccctaaa tttcttagtg attaacctg gctaaacccc	2940
ccaagaagaa gcgcaagggt aacatcaaaa aaaaccaggt aatgaacctg ggtccgaact	3000
ctaaactgct gaaagaatac aaatcccagc tgatcgaact gaacatcgaa cagttcgaag	3060
caggatcggg tctgatcctg ggtgatgctt acatccgttc tcgtgatgaa ggtaaacct	3120
actgtatgca gttcgagtgg aaaaacaaag catacatgga ccacgtatgt ctgctgtacg	3180
atcagtgggt actgtccccg ccgcacaaaa aagaacgtgt taaccacctg ggtaacctgg	3240
taatcacctg gggcgcccag actttcaaac accaagcttt caacaaactg gctaacctgt	3300
tcatcgtaa caacaaaaaa accatcccga acaacctggt tgaaaactac ctgaccccga	3360
tgtctctggc ataactggtc atggatgatg gtggtaaatg ggattacaac aaaaactcta	3420

[0022]

ccaacaaatc gatcgtagtg aacacccagt ctttcacttt cgaagaagta gaataacctgg	3480
ttaaggggtct gcgtaacaaa ttccaactga actgttagct aaaaatcaac aaaaacaaac	3540
cgatcatcta catcgattct atgtcttacc tgatcttcta caacctgac aaaccgtacc	3600
tgatcccgca gatgatgtac aaactgccga acactatctc ctccgaaact ttcctgaaat	3660
agggctagca agcttggaca cgctgaaatc accagtctct ctctacaaat ctatctctct	3720
ctatcttctc cataataatg tgtgagtagt tcccagataa gggaattagg gttcctatag	3780
ggtttcgctc atgtgttgag catataagaa acccttagta tgtatttgta ttgtaaaat	3840
acttctatca ataaaatttc taattcctaa aaccaaatac cagtactaaa atccagatca	3900
tgcattgtac agcggccgcg ttaacgcgta tactctagat acgtacgat cgccatggag	3960
ccatttaca ttgaatatat cctgccg	3987

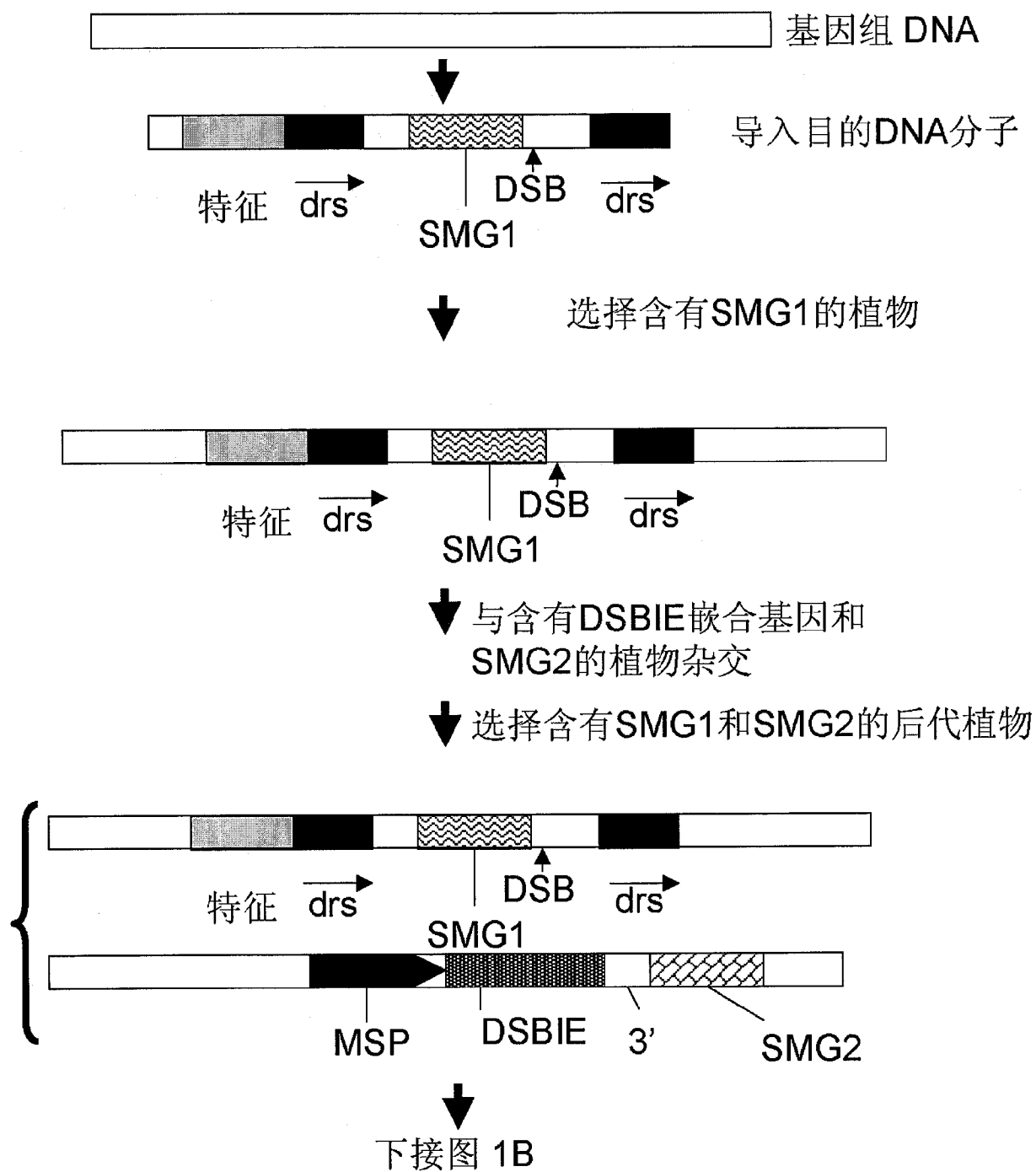


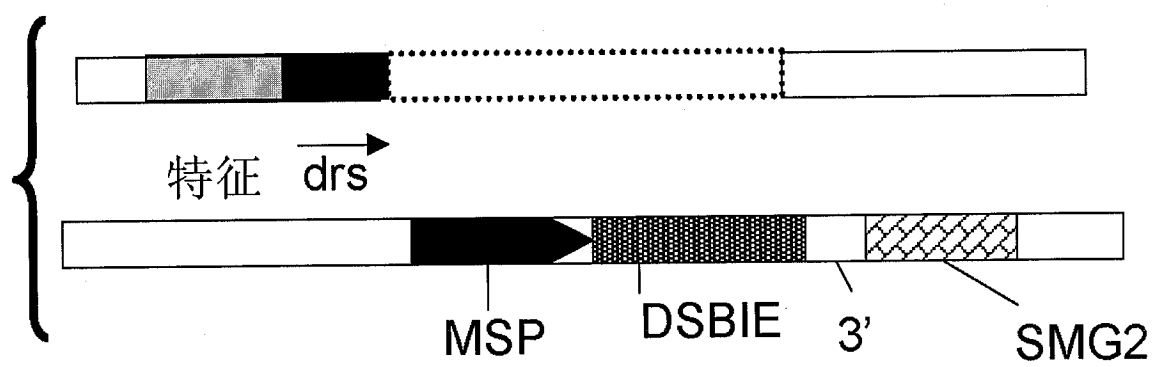
图 1A

上接图1A

↓ 通过花粉与受体植物杂交

↓ 选择存在SGM2的后代

↓ 筛选缺乏SGM1的后代



↓ 与受体植物杂交

↓ 筛选缺乏SGM2的后代

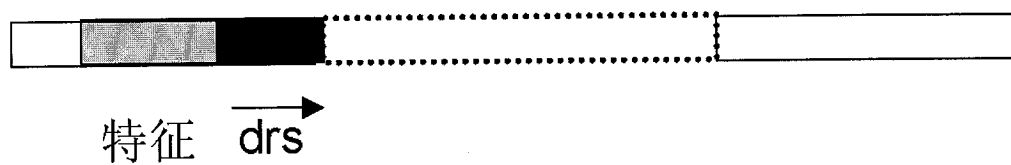


图 1B

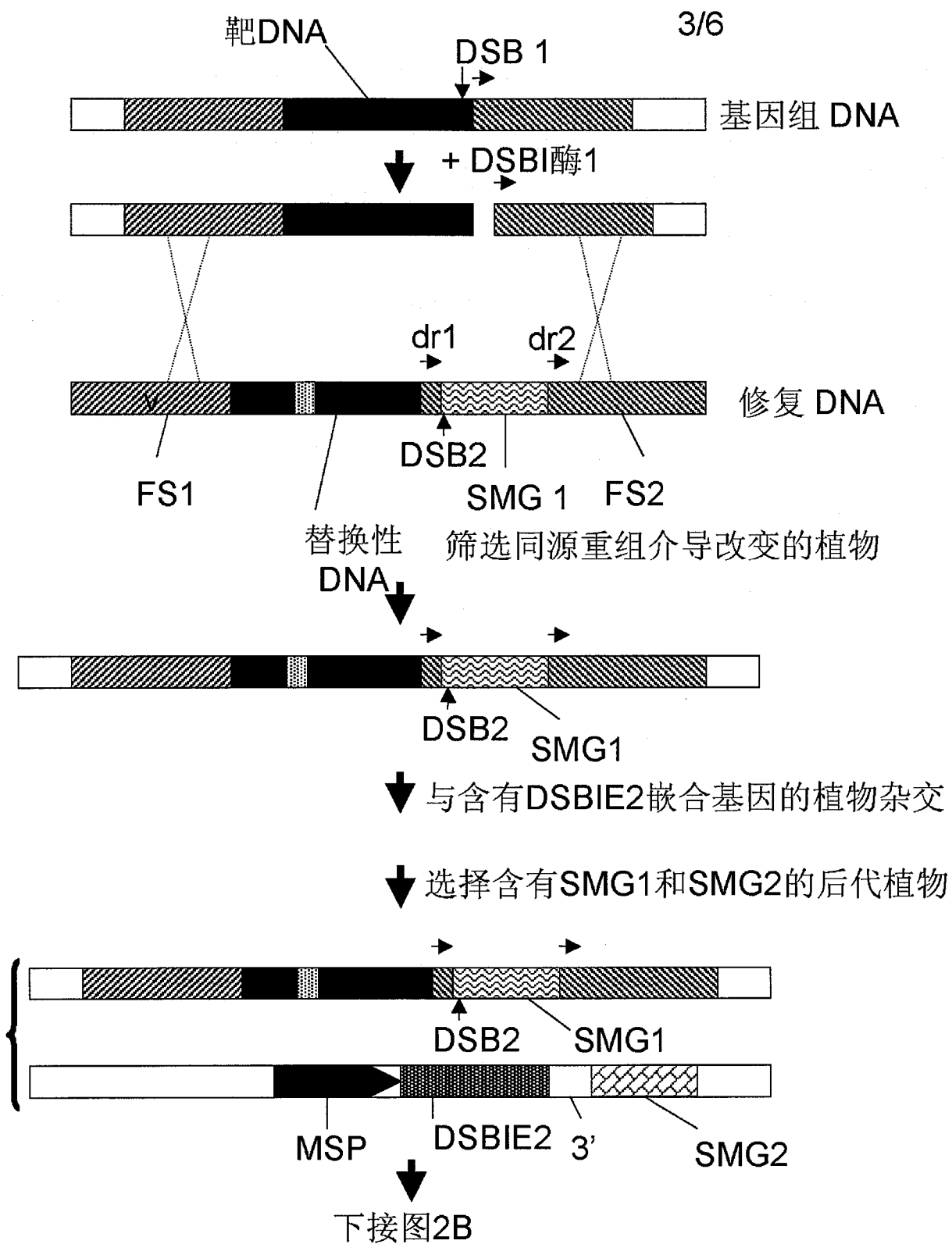


图 2A

上接2A

↓ 通过花粉与受体植物杂交

↓ 选择存在SGM2的后代

↓ 筛选缺乏SGM1的后代

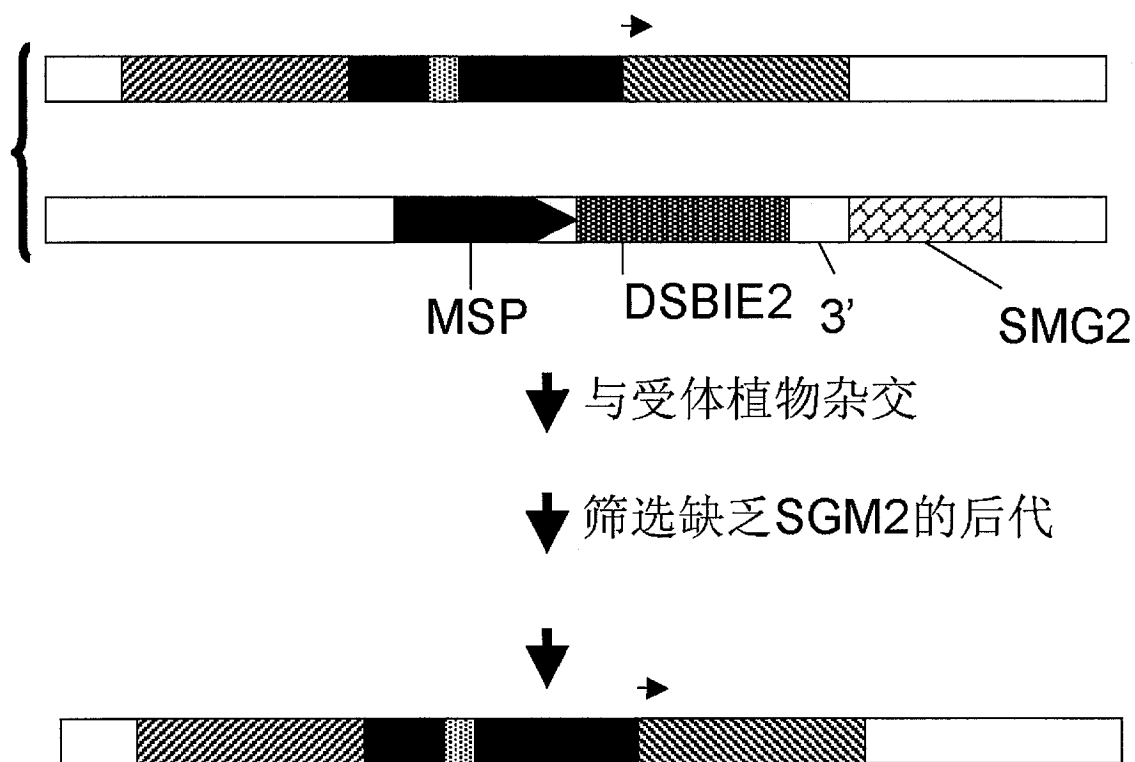


图 2B

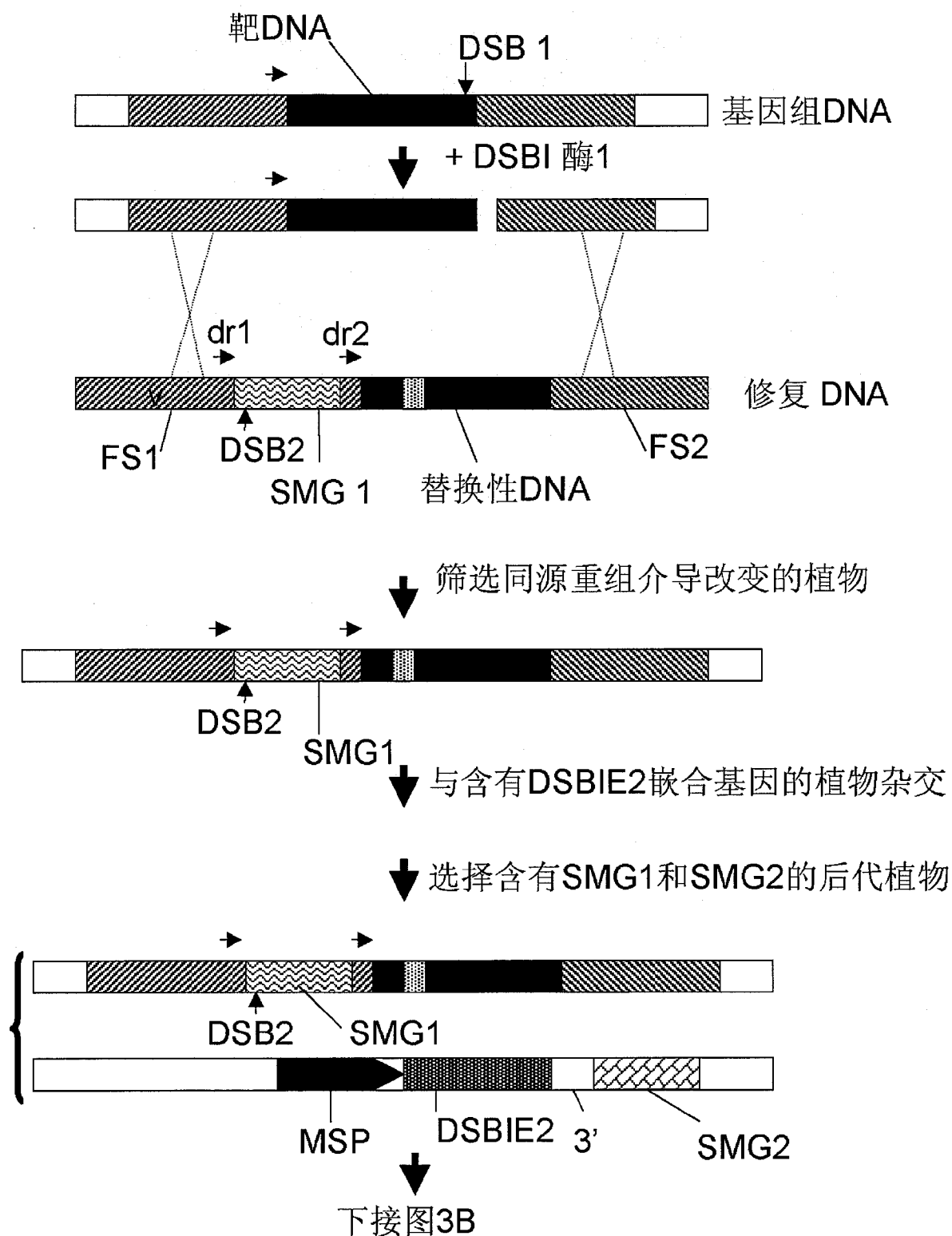


图 3A

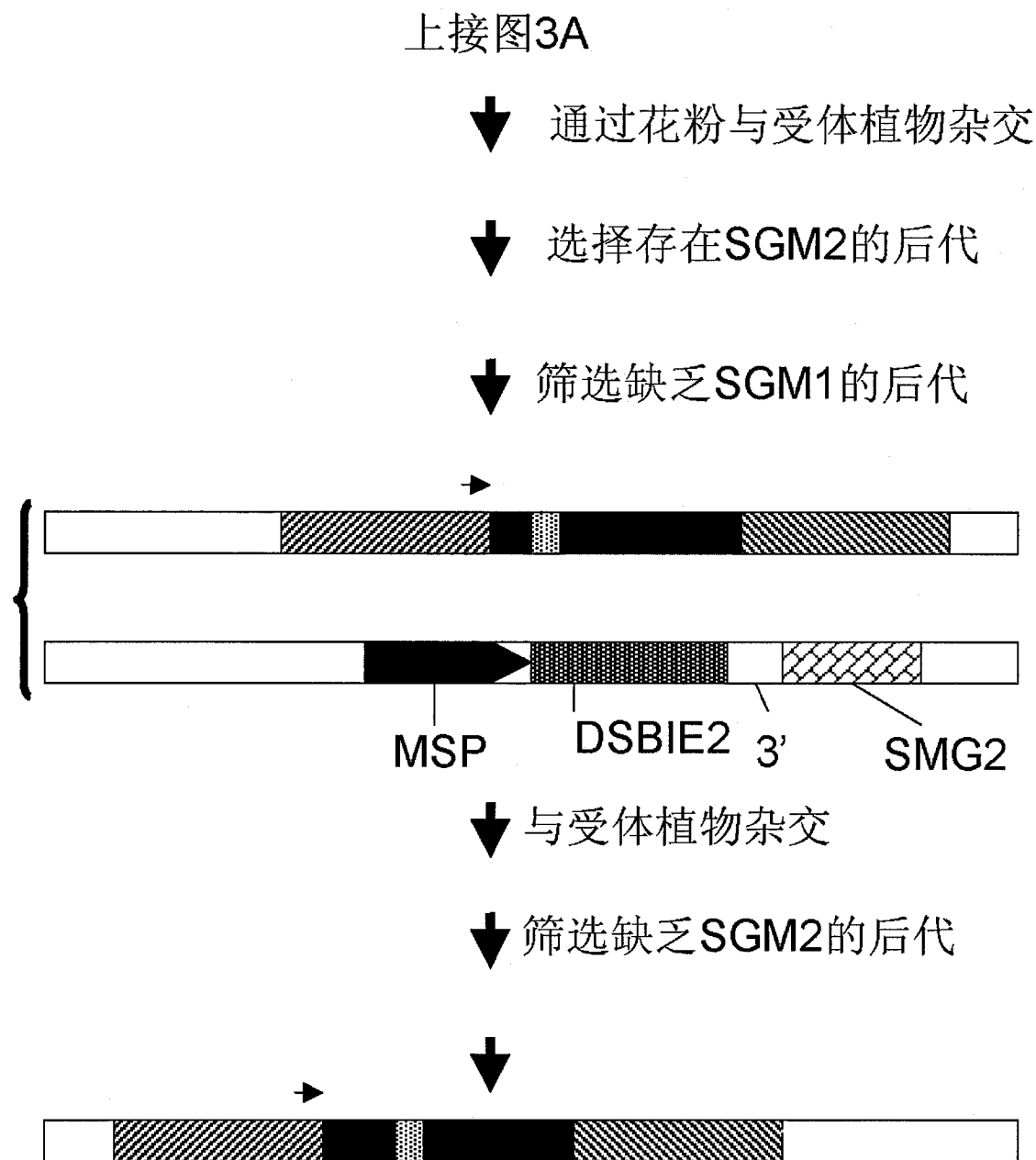


图 3B